Is model construction open to strategic decisions?
An exploration in the field of linear reasoning

Vicky Dierckx and André Vandierendonck

Ghent University, Belgium

Mario Pandelaere

University of Leuven, Belgium

This paper reports four experiments investigating whether model construction of linear reasoning problems is open to strategic decisions. A reversed choice/no-choice paradigm was used in which reasoners first had to apply two model construction strategies (acronym and rehearsal strategy) to two problem sets. Next, they could choose freely among the two strategies to apply to a new problem set. Experiment 1 showed that reasoners selected the strategy that they experienced as the most accurate one in the no-choice phase. Moreover, in Experiment 2, it was found that reasoners adapted their strategy choice to changing problem features, to use the most suitable strategy for premise encoding. Experiments 3 and 4 generalised these findings to more complex linear reasoning problems with a mixed sentence frame and a semi-continuous presentation of the premises, and to two-model problems. On the basis of these results, we argue that strategic decisions influence model construction in linear reasoning.

For a long time, research in deductive reasoning has been driven by intense debates on the nature of the reasoning mechanism underlying all types of deductive reasoning. The heart of this controversy concerned the question of whether reasoning involves the syntactic application of natural deduction rules (mental rule theories; e.g., O’Brien, Braine, & Yang, 1994; Rips, 1994) or is based upon the construction and manipulation of mental models representing the
meaning of the premises (mental models theory; e.g., Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992, 1994). Nowadays, the dispute seems to be settled in favour of the mental models theory (e.g., Schaeken, De Vooght, Vandierendonck, & d’Ydewalle, 2000; Schaeken, Vandierendonck, Schroyens, & d’Ydewalle, in press), although the existence of individual differences in reasoning is still recognised (Roberts, 1993, 2000a, 2000b). Since relatively much research has been invested in opposing both types of theories, the mental models theory has stayed fairly general.

The mental models theory assumes that reasoning encompasses three phases: model construction, conclusion generation, and the search for counterexamples, or alternative models, in which the conclusion does not hold (conclusion validation), but remains relatively vague about the exact nature of these three phases. According to Evans (1989, 2000a; Evans & Over, 1997), reasoning is mainly based on implicit processes. Following his view, pragmatic heuristic processes determine the modelling of the premises and the formulation of a putative conclusion. Reasoners thus do not seem to have control of which information they model nor of how this modelling is actually carried out. Furthermore, once a putative conclusion is inferred, usually no further attempts at validating the conclusion seem to be made (e.g., Evans, Handley, Harper, & Johnson-Laird, 1999; Newstead, Handley, & Buck, 1999; Polk & Newell, 1995). However, in some situations, such as when the conclusion conflicts with people’s beliefs (e.g., Evans, Newstead, Allen, & Pollard, 1994; Newstead, Pollard, Evans, & Allen, 1992; Oakhill, Johnson-Laird, & Garnham, 1989) or when the need to reason logically is stressed (e.g., Evans et al., 1994; Newstead et al., 1992), reasoners do seem to search for counterexamples. Based on these latter findings Evans argues that, unlike model construction and conclusion generation, conclusion validation is strategic in nature.

The present study explores whether model construction is really an implicit process, as suggested by Evans (1989, 2000a), or whether the modelling of the premises is open to strategic decisions. The question is thus raised whether reasoners are able to choose among model-construction strategies taking into account the benefits of the strategies, or their accuracy and their speed, as well as their costs or the amount of resources required for their execution (e.g., Payne, Bettman, & Johnson, 1993; Siegler & Lemaire, 1997). A strategic choice between available strategies then implies a weighting of these benefits and costs, to select a strategy with which the most accurate results can be obtained as quickly as possible and with an acceptable amount of effort (e.g., Payne et al., 1993; Siegler & Lemaire, 1997). Whether this weighting is a completely conscious process (e.g., Evans, 2000a) or whether it is mostly based on implicit knowledge (e.g., Siegler & Lemaire, 1997) is still a matter of debate and irrelevant for the purpose of our study.

An interesting field in which to investigate the strategic nature of model construction is relational reasoning or reasoning about one- or two-dimensional
relations between terms (e.g., “visiting London before visiting Brussels”, “visiting Paris before visiting London”), since several studies have already explored strategy use in encoding and solving these types of problems. Inspection of this literature reveals that these studies can be broadly categorised into three groups. First, there are studies detecting strategies inter-individually (e.g., Egan & Grimes-Farrow, 1982; Sternberg & Weil, 1980). Here, reasoners are classified into different strategy groups based on their verbal protocols and/or their performance measures. Subsequently, the reasoners’ strategy choice is linked with their performance on verbal and spatial ability tests. Reasoners were found to solve the problems with the strategy that was the most compatible with their abilities, suggesting that a strategic choice underlay problem solving (Grimes, 1980, cited in Egan & Grimes-Farrow, 1982; Sternberg & Weil, 1980). Yet, with this procedure, it is unclear whether more than one strategy was available to the reasoners. Possibly, they simply applied the single strategy that came to their mind without engaging in a strategic decision. Furthermore, Sternberg and Weil (1980) found that reasoners did not always follow the instruction to solve the problems with a specified strategy. It is unclear whether these reasoners compared the benefits and the costs of the instructed strategy with those of other strategies and discarded the instructed one because they preferred one of the other strategies, or whether they had no control of premise encoding.

Second, there are studies indicating that over a large set of problems, reasoners switch from a representational strategy, in which the premises are encoded into a model, to a short-cut strategy for which the construction of a representation is redundant (Quinton & Fellows, 1975; Wood, Shotter, & Godden, 1974). These strategy switches, however, can only be realised using a specific procedure in which a large number of problems are presented together with their conclusion. Indeed, Roberts (2000a) could not replicate these findings in an experiment with only six problems, not even by explicitly instructing the short-cut strategy. Ohlsson (1984) also reported strategy switches over problems but, because the verbal protocols of only five participants were clear enough to distinguish the strategies, the results are not very reliable.

Finally, there is one study investigating strategies in model construction itself (Schaeken & Johnson-Laird, 2000). Here, multiple-model problems were presented with an irrelevant premise that did not have to be represented in the model(s) to solve the problems. Furthermore, if reasoners ignored the irrelevant premise, the problems became a lot easier because then they could be encoded in one single model. It appeared that reasoners tended to only apply the “reduced model construction strategy”, if they had been trained with problems having many irrelevant premises. Otherwise, they needlessly kept on taking into account the multiple possibilities of representing the temporal relation between the premise terms.

Taken together, these studies provide some tentative evidence indicating that the encoding of relational reasoning problems is open to strategic decisions, although the evidence is far from conclusive. Moreover, most of these studies did
not consider strategic choices in model construction. Often the use of a model-construction strategy was compared with the use of a short-cut strategy or a rule-based strategy that did not involve the construction of a mental model. Consequently, whether implicit or strategic processes guide model construction still remains an open question that will be investigated systematically in the present study.

THE STUDY OF MODEL-CONSTRUCTION STRATEGIES

A useful paradigm to detect whether a strategic choice underlies model construction is the choice/no-choice paradigm of Siegler and Lemaire (1997). In this paradigm, participants are first allowed to choose freely one strategy, of a limited set of strategies, to apply to a problem set. Subsequently, in the no-choice phase, they have to apply all strategies to all problems such that the accuracy and speed of the strategies for these problems can be recorded and linked with the participants’ strategy choice in the choice phase. The prerequisite for a successful application of this procedure is, of course, that a reliable operational distinction between the strategies is possible. This is a problem for the study of strategic processes in model construction because particularly visual and spatial processes are involved in the construction process. During model construction reasoners imagine positioning mental tokens, representing the premise terms, such that the relation described in the premises is reflected in the structure of the model (Johnson-Laird, 1983). For relational reasoning problems, the model often has the structure of a spatial array, representing mental tokens one beside or underneath the other (e.g., DeSoto, London, & Handel, 1965; Huttenlocher, 1968). As a consequence, introspection is often the only means to assess what type of model is constructed and/or which kind of tokens are used to represent the premise terms (e.g., Egan & Grimes-Farrow, 1982; Huttenlocher, 1968; Quinton & Fellows, 1975). However, as already identified by Nisbett and Wilson (1977), this method is not a reliable way of gaining insight into one’s mental processing. In order to avoid these problems, we have decided to “teach” reasoners two model-construction strategies, of which the type of mental tokens varies and, at the same time, can be distinguished by the verbalisations reasoners make during the positioning of the tokens. For this, we were inspired by the work of Mynatt and Smith (1977). In their study, participants received relational reasoning problems, describing a one-dimensional relation between premise terms (linear reasoning problems), of which they had to recall the ordering of the terms. Besides a visuo-spatial strategy, participants reported using verbal strategies consisting of rehearsing the premise terms (rehearsal strategy) or only the first letters of these terms (acronym strategy). Within the context of solving linear reasoning problems, Huttenlocher (1968, p. 560) also recognised the existence of both ways of encoding the premises by mentioning that “The S[subject]s represent items as words (or abbreviations) rather than pictures ...."
Table 1 illustrates the application of the rehearsal and the acronym strategy to linear reasoning problems describing a temporal order of terms. For the rehearsal strategy, reasoners are instructed to encode the entire premise terms into a string such that the position of the terms in the string reflects the order in the premises, and to rehearse the string to maintain the order of the terms in memory so as to be able to deduce the relative order of any two premise terms, once the premises have disappeared from the screen. In the example, the first premise states that the “bank (is visited) before (the) abbey”, so the string starts with “bank–abbey”. The second premise states that the “abbey (is) before (the) villa”. This information can be encoded by attaching the token “villa” to the token “abbey” so that the string is expanded to “bank–abbey–villa”. And so on for every new premise. In order to evaluate the conclusion, such as for instance “villa before tower”, reasoners are told to check whether the first term, “villa”, occurs earlier in the string than the second term, “tower”. If this is the case, the conclusion is valid, otherwise the proposed conclusion is invalid. The procedure is similar for the acronym strategy except that now only the first letters of the terms are placed in a string (see example in Table 1: “bavith”).

Both the rehearsal and the acronym strategy are in fact memory strategies often used spontaneously to memorise a list of words (e.g., Sternberg, 1996). The question may be raised whether these strategies can be considered genuine model-construction strategies. To answer this question, a closer look should be taken at what exactly takes place during model construction. According to the mental models theory, model construction is the translation of the meaning of the premises into an integrated representation in working memory. The models theory is unspecific about how this translation process is actually carried out and about which type of resources are tapped during this process, but it is evident that memorisation processes are inevitably involved in this first phase of the reasoning process. The only aspect about which the theory is clear, is that the representation constructed should satisfy the defining characteristic of a mental

<table>
<thead>
<tr>
<th>Problem</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank before abbey</td>
<td>Rehearsal strategy</td>
</tr>
<tr>
<td>Villa before igloo</td>
<td>Acronym strategy</td>
</tr>
<tr>
<td>Igloo before tower</td>
<td>“b–a–v–i–t–h”</td>
</tr>
</tbody>
</table>

An example of a temporal linear reasoning problem encoded with the rehearsal strategy and with the acronym strategy.
model: the principle of structural identity (isomorphism). According to this principle, each distinct token in the model must correspond to an object named by a premise term and the relation between the objects, as described in the premises, must be preserved in the model (Johnson-Laird, 1983, pp. 419, 422). The nature of the tokens actually used and how the relation between the tokens is represented, is not specified by the theory (see Johnson-Laird, 1983, pp. 422–430; Rickheit & Sichelschmidt, 1999, p. 11). The representation achieved with the rehearsal or the acronym strategy does represent each distinct premise term by a different mental token, respectively a different word or a different letter. The strategies only differ in the type of mental token used, hence making them suitable alternative model-construction strategies.

We report four experiments exploring by means of a variant of the choice/no-choice paradigm whether reasoners are able to choose strategically between the acronym and the rehearsal strategy. Because it could not be ensured that all reasoners knew how to apply these strategies to linear reasoning problems, we decided to reverse the order of the choice and the no-choice phase (see also Pressley, Levin, & Ghatala, 1984). This adaptation also excludes the possibility that, in the no-choice phase, performance with the preferred strategy is better than performance with the non-preferred strategy because of a practice effect rather than because it is really the most efficient strategy.

In our experiments, the first phase is hence the no-choice phase in which participants have to apply the acronym and the rehearsal strategy to two sets of linear reasoning problems. Subsequently, in the choice phase, they are free to choose which strategy to apply to a new set of problems with the instruction to solve these problems as accurately as possible. By stressing the importance of accuracy, the strategy-selection process will probably focus on maximising accuracy, possibly at the expense of minimising the speed and cost of executing a strategy. When model construction is indeed open to strategic decisions, we expect people to select the strategy that leads to the most accurate results. This hypothesis was tested in four experiments.

EXPERIMENT 1

In the first experiment we manipulated the perceived efficiency of the acronym and the rehearsal strategy by varying the characteristics of the reasoning problems to which these strategies had to be applied in the no-choice phase. Two sets of problems were presented: one set suitable for encoding with both the rehearsal and the acronym strategy and one set less suitable for encoding with either strategy. The former set consisted of problems having short premise terms all starting with different letters (easy problems). The other set of problems was composed of long premise terms, some having the same starting letter (hard problems). These latter problems are harder to encode with the rehearsal strategy because it requires more resources to rehearse a string of long premise terms than
to rehearse a string of short terms (word-length effect, e.g., Baddeley, Thomson, & Buchanan, 1975). Consequently, the model constructed of these problems may be less stable, through which the position of the terms may be altered and a conclusion may have to be evaluated on an incorrect representation. The hard problems are also more difficult to encode with the acronym strategy since constructing an acronym of only the first letters of the terms will lead to an ambiguous representation in which some letters are repeated. Therefore, it is impossible to know which letter represents which term, unless some extra information is encoded that distinguishes the representations of these terms. This process, however, will require more resources than merely encoding the first letters of the terms and may result in less accurate conclusion evaluations.

In the no-choice phase of the experiment, half of the participants had to encode the hard problems with the rehearsal strategy and the easy problems with the acronym strategy. For this group, the acronym strategy will be perceived as the most efficient strategy. For the other half of the participants, the complementary situation was created such that they would perceive the rehearsal strategy as the most efficient strategy. To check the participants’ perception of the efficiency of the strategies, they were asked to make an estimation of the number of reasoning problems they believed they had solved correctly with both strategies. We expect participants to encode the problems in the choice phase with the strategy they have perceived as being the most efficient one in the no-choice phase or, in other words, with the strategy they have applied to the easy problems.

Method
Participants and design
A total of 32 first-year students enrolled at the faculty of Psychology and Educational Sciences at Ghent University participated for course requirements and partial credit. The design was a 2 (strategy application order) × 2 (strategy type: acronyms vs rehearsal) × 2 (problem set: easy vs hard) × 2 (group) design of which the last three factors formed a Latin-Square cross-over design (Lee, 1975). The participants were randomly assigned to the cells of the two between-subjects conditions: strategy application order and group. The participants of group 1 applied the acronym strategy to the easy problems and the rehearsal strategy to the hard problems in the no-choice phase, whereas the participants of group 2 applied the acronym strategy to the hard problems and the rehearsal strategy to the easy problems.

Materials
A total of 32 temporal linear reasoning problems were constructed: 2 practice problems, 2 × 10 problems for the no-choice phase, and 10 problems for the
choice phase. All problems consisted of five premises, linking six objects/subjects with the temporal relation “before” (constant sentence frame, Mynatt & Smith, 1977). The premises were isotropic: the second term of each premise reoccurred as the first term in the next premise. On the basis of each problem, a conclusion of the relation between two terms had to be evaluated. To avoid anchoring effects (e.g., Potts, 1972), this conclusion never involved a relation with an end-term. For 2 problems of each set of 10 problems, the conclusion queried the relation between the third and the fourth term. For all other problems, including the practice problems, the conclusion queried a relation between two terms not occurring together in the same premise, so that a transitive inference was required.

Two different problem sets were created. The first set (easy problems, $n = 22$) consisted of problems suitable for encoding with both the acronym and the rehearsal strategy. The problems were composed of short premise terms (< 6 letters) that started with different letters. The terms occurred in such an order that a Dutch-sounding pseudo-word could be formed with the starting letters. Two of these problems were used as practice problems in the no-choice phase, ten problems as easy problems in the same phase and the remaining ten problems were presented in the choice phase. The second problem set (hard problems, $n = 10$ problems) consisted of problems that were equally unsuited for encoding with the acronym as with the rehearsal strategy. These problems had long (between 5 and 11 letters) premise terms, of which two pairs of two non-end terms started with the same letter.

Procedure

The experiment was conducted individually. Participants first received a general instruction describing the type of problems to be solved. Next, they were taught one model-construction strategy that they had to apply aloud to a practice problem and subsequently to 10 easy or to 10 hard problems (no-choice phase). Then, the other strategy was explained and had to be applied aloud, again to a practice problem and to the 10 remaining problems of the no-choice phase. After the no-choice phase, participants were asked to make an estimation of the number of problems they believed they had solved correctly with each strategy. Then the choice phase started. Participants were told that they could freely choose between the two strategies to apply to the next 10 problems and that they had to solve the problems as accurately as possible. It was left unspecified whether participants could switch between strategies. Again, the strategies had to be applied aloud.

The premises of each problem were presented simultaneously for 6 seconds on the screen of an IBM-compatible PC-AT 386. After the premises had disappeared, a conclusion was presented. Participants had to press the left mouse button when they believed the conclusion was correct, and to press the right mouse button otherwise. The response given and the conclusion evaluation times
(measured from the time the conclusion was presented until a response was given) were recorded. In the choice phase, the strategy selected for each problem was also recorded.

Results

The results of the no-choice phase and of the choice phase are presented separately. The data of one participant, with conclusion evaluation times lower than 200 ms, were discarded from the analyses.

No-choice phase

Accuracy of conclusion evaluation, estimated accuracy, and mean conclusion evaluation times of the correctly evaluated problems were analysed according to a 2 (strategy application order) × 2 (strategy type) × 2 (problem set) × 2 (group) Latin Square cross-over design. Since there were no differences as a function of strategy application order, this variable will not be discussed further.

Accuracy. The average proportion of correctly solved problems is displayed in Table 2, as a function of the type of strategy applied and of the problem set presented. There was no effect of strategy type but performance differed as a function of problem set, $F(1, 29) = 30.14, p < .001$. As expected, the hard problems were solved less accurately ($M = 66\%$) than the easy problems ($M = 87\%$). There was also a main effect of group, which, in our design, can be interpreted as an interaction between strategy type and problem set, $F(1, 29) = 14.97, p < .001$. The accuracy of the acronym strategy varied as a function of problem set, $t(29) = 8.75, p < .001$, whereas the accuracy of the rehearsal strategy stayed about the same. Otherwise stated: for the easy problems the acronym strategy was more accurate than the rehearsal strategy, $t(29) = 3.91, p < .001$ and for the hard problems the rehearsal strategy led to more accurate results than the acronym strategy. This difference, however, fell short of significance, $t(29) = 1.86, p = .073$.

Estimated accuracy. The results of the estimated accuracy data (see Table 2) are similar to the results of the actual accuracy data. Participants thought they had solved fewer hard problems ($M = 48\%$) correctly than easy problems ($M = 77\%$), $F(1, 29) = 83.83, p < .001$. There was also an effect of group, showing an interaction between strategy type and problem set, $F(1, 29) = 12.92, p < .01$. The acronym strategy was estimated to be more accurate than the rehearsal strategy when applied to the easy problems, $t(29) = 2.91, p < .01$. There was no difference in estimated accuracy between the strategies when applied to the hard problems, $t(29) = 1.52, p = .140$.

The correlations between the accuracy and the estimated accuracy of both strategies are presented in Table 3, above the diagonal. As can be seen in the
significant positive correlations are found between the actual accuracy and the estimated accuracy of both strategies. All other correlations are negative, as may be expected since one strategy was applied to easy problems and the other strategy to hard problems.

**Conclusion evaluation time.** There were no main effects of strategy type or of problem set (see Table 2) but there was again an effect of group, $F(1, 29) = 16.53, p < .001$, showing that only the evaluation time of the acronym strategy varied as a function of problem set, $t(29) = −3.31, p < .01$. Otherwise put: The hard problems were solved faster with the rehearsal than with the acronym strategy, $t(29) = −2.90, p < .01$. The opposite tendency was present for the easy problems, $t(29) = −1.93, p = .064$.

**Choice phase**

**Strategy selection.** The percentage of acronym selection of the participants who applied the acronym strategy to the easy problems and the rehearsal strategy

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Easy problems</th>
<th></th>
<th>Hard problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acronyms (n = 16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.96</td>
<td>.61</td>
<td>.71</td>
</tr>
<tr>
<td>SD</td>
<td>.06</td>
<td>.04</td>
<td>.16</td>
</tr>
<tr>
<td>Rehearsal (n = 15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.84</td>
<td>.43</td>
<td>.52</td>
</tr>
<tr>
<td>SD</td>
<td>.11</td>
<td>.05</td>
<td>.14</td>
</tr>
<tr>
<td>Acronyms (n = 15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rehearsal (n = 16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.9</td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>SD</td>
<td>0.9</td>
<td></td>
<td>1.4</td>
</tr>
</tbody>
</table>
to the hard problems in the no-choice phase \((n = 16)\), amounted to 99%. This percentage only amounted to 9% for the participants who applied the acronym strategy to the hard problems and the rehearsal strategy to the easy problems in the no-choice phase \((n = 15)\).

In order to detect the principles underlying the strategy selection process, the data were submitted to logistic regression analyses with a binary variable as dependent variable indicating, for each problem, whether or not the acronym strategy was selected. We used the generalised estimation equation approach to logistic regression to take into account dependencies between observations arising from a single participant (e.g., Diggle, Kung-Yee, & Zeger, 1994). The predictors were the problem set to which the acronym strategy was applied in the no-choice phase (easy = 1, hard = 0), strategy application order (A/R = 1; R/A = 2), the accuracy scores, the estimated accuracy scores, and the evaluation times of the acronym and the rehearsal strategy in the no-choice phase. In line with Siegler and Lemaire (1997), we also included difference scores as predictors in the analyses. These scores reflected the relative accuracy (accuracy acronym – accuracy rehearsal), the relative estimated accuracy, and the relative evaluation time of the strategies. Also, the most accurate strategy, the estimation of the most accurate strategy, and of the fastest strategy were calculated on the basis of these difference scores. In order to avoid multicolinearity, two or more predictors being a linear combination one of the other were not included in the same model. Table 4 presents the individual logistic regressions for each significant predictor. Problem set was the best predictor, followed by the estimation of the most accurate strategy.

We used a backward selection method to determine the best-fitting model (i.e., the model containing only significant predictors while at the same time maximising the proportional reduction in quadratic error; \(PRE\)). This model had

<table>
<thead>
<tr>
<th></th>
<th>Acronyms</th>
<th>Rehearsal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated</td>
<td>–.37*</td>
<td>–.25</td>
</tr>
<tr>
<td>Accuracy</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Acronyms</td>
<td>–.37*</td>
<td>–.25</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Accuracy</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Acronyms</td>
<td>–.37*</td>
<td>–.25</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Accuracy</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Acronyms</td>
<td>–.37*</td>
<td>–.25</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Accuracy</td>
<td>–.17</td>
<td>–.11</td>
</tr>
<tr>
<td>Estimated</td>
<td>–.17</td>
<td>–.11</td>
</tr>
</tbody>
</table>

\* \(p < .05\), \** \(p < .01\), \*** \(p < .001\)
only one predictor: “problem set”. The model indicates that when the acronym strategy is applied to easy problems in the no-choice phase—and the rehearsal strategy to hard problems—there is a higher probability of selecting the acronym strategy in the choice phase (and a lower probability of selecting the rehearsal strategy) than when the acronym strategy is applied to hard problems (and the rehearsal strategy to easy problems).

**Strategy performance.** Participants who applied the acronym strategy in the choice phase solved more problems correctly ($M = 99\%$) than those applying the rehearsal strategy ($M = 87\%$), $t(29) = -3.66, p < .001$. However, there were no differences in evaluation time (respectively $M = 2.8$ s and $M = 3.1$ s), $t < 1$.

**Discussion**

By using easy and hard problem sets we managed to manipulate the perceived efficiency of the strategies with which the linear reasoning problems were encoded. When a strategy was applied to easy problems in the no-choice phase, participants believed it produced more accurate results than the other strategy, applied to hard problems. In general, the easy problems were also solved more accurately than the hard problems.

Both in the choice and in the no-choice phase, more easy problems could be solved correctly with the acronym strategy than with the rehearsal strategy. This better performance may be explained by the fact that only six letters, forming one
pseudo-word, have to be maintained in working memory rather than six separate words. Moreover, “scanning” this pseudo-word to evaluate a conclusion is also less demanding than searching a sequence of words. The hard problems, on the contrary, were solved faster and slightly better with the rehearsal strategy than with the acronym strategy. It seems that the rehearsal strategy can be more flexibly applied to the hard problems than the acronym strategy. This is no surprise because when a pseudo-word is constructed by combining identical letters, it is impossible to know which of the two premise terms, starting with the same letter, preceded the other, unless extra information is stored. However, when long premise terms are encoded with the rehearsal strategy, it is still possible to infer the order of the terms. In fact, neither the accuracy of the rehearsal strategy nor its evaluation time varied as a function of problem set. The presumed increase in mental load of processing long premise terms thus did not result in a decreased accuracy or in an increased evaluation time.

Remarkably, reasoners who applied the acronym strategy to the hard problems were still able to solve 60% of the problems correctly—higher than chance level, $t(15) = 2.69, p < .05$. This indicates that somehow they did manage to add some extra information to the ambiguous representation, making it possible to distinguish the premise terms starting with the same letters. In that case, not only the pseudo-word but also the additional information must be recollected when evaluating the conclusion, and this demands more resources than only recollecting the pseudo-word. This increase in load was indeed reflected in the respective evaluation times of the hard and the easy problems encoded with the acronym strategy: 4.3 s vs 2.9 s.

The best predictor of strategy choice was the type of problem to which the strategy was applied before. When a strategy was applied to easy problems in the no-choice phase, there was a higher probability of selecting that strategy in the choice phase than when that strategy was applied to hard problems. Although the participants perceived the latter strategy as the least efficient one, all subjective measures of the accuracy of the strategies were worse predictors than the problem set to which the strategy was applied. The estimation of the most accurate strategy was found to be the best predictor reflecting the participants’ appraisal of the strategy. Closer inspection of the data shows that this predictor was only worse than problem set because of two participants. One of them estimated the accuracy of both strategies equally high, and the other estimated the rehearsal strategy as being more accurate than the acronym strategy but did select the acronym strategy in the choice phase. When these two observations were discarded from the analyses, the logistic regression model with the estimation of the most accurate strategy as predictor was identical to the model with problem set as predictor. So, almost all participants selected the strategy that they considered the most efficient one.

A final remark about the data concerns the observation that the most accurate strategy, according to the participants’ estimations, was a better predictor than
the actual most accurate strategy (see Table 4). This shows that the strategy selection process is guided by a personal evaluation of the strategies. The kind of evaluation taking place seems to be very elementary. Not the estimation of the actual number of correctly solved problems, nor the estimated accuracy of one strategy relative to the other, but the mere fact that one strategy is considered to be better than the other, determines strategy choice. Taken together, these results suggest that model construction of linear reasoning problems is indeed accessible to strategic decisions.

EXPERIMENT 2

In the previous experiment, the perceived efficiency of the strategies was manipulated by varying the problem features in the choice phase. After this phase, reasoners were asked to give their appraisal of the strategies and to choose between the strategies to solve a new set of problems as accurately as possible. In this situation, it would be very strange to apply the strategy that was given the lowest appraisal. The procedure may thus have prompted reasoners to select the strategy that they judged to be the most accurate one. In the present experiment we therefore manipulated the efficiency of the strategies only after a judgement of their accuracy had been given by varying the features of the problems presented in the choice phase. Three types of problems were presented: problems having features that made them suitable to encode with the acronym strategy but harder to encode with the rehearsal strategy (biased towards acronyms), problems with features that made them suitable to encode with the rehearsal strategy but harder to encode with the acronym strategy (biased towards rehearsal), and unbiased filler problems that could suitably be encoded with both strategies. If strategic processes guide the encoding of the problems in the choice phase, strategy switches may be expected to occur adaptive to changing problem features. More specifically, it may be expected that the problems biased towards acronyms will be more frequently encoded with the acronym strategy than with the rehearsal strategy, and the problems biased towards rehearsal more frequently with the rehearsal strategy than with the acronym strategy.

Method

Participants and design

A total of 20 first-year students of the faculty of Psychology and Educational Sciences at Ghent University participated for course requirements and partial credit. None of them had participated in the previous experiment. All participants applied the acronym and the rehearsal strategy (within-subject variable) to five problems each. The order of strategy application was again a between-subjects variable.
**Materials and procedure**

A total of 42 six-term temporal linear reasoning problems were used: 2 practice problems, 2 × 5 problems for the no-choice phase and 30 problems for the choice phase. The practice problems, the problems of the no-choice phase, and 10 problems of the choice phase were composed of short premise terms (< 6 letters) all beginning with different letters (Type 1: unbiased problems; identical to the easy problems of Experiment 1). Ten problems were composed of long terms (between 5 and 11 letters) all starting with different letters that could be combined in a Dutch-sounding pseudo-word (Type 2: problems biased towards acronyms). The remaining 10 problems consisted of six short premise terms (< 6 letters), of which two pairs started with the same letters (Type 3: problems biased towards rehearsal).

The procedure of the present experiment was identical to that used in the previous experiment, except that in the no-choice phase the acronym and the rehearsal strategy had to be applied to five problems that were suited for encoding with both strategies. Furthermore, in the choice phase, participants were free to choose which strategy to apply to 30 problems, 10 of each type. These 30 problems were presented in a completely randomised order.

**Results**

**No-choice phase**

The accuracy of the strategies, the participants’ estimation of the accuracy of the strategies, and the mean evaluation time of the correctly solved problems were analysed by a 2 (strategy application order) × 2 (strategy type) ANOVA. Because there were no differences as a function of strategy application order, this variable will not be discussed further.

**Accuracy and estimated accuracy.** There were no accuracy differences between the acronym (M = 92%) and the rehearsal strategy (M = 88%), F < 1. Nor were there differences in the estimation of the number of problems solved correctly between the acronym (M = 76%) and the rehearsal strategy (M = 72%), F < 1.

A correlation analysis of the accuracy and the estimated accuracy of both strategies (see Table 3, below the diagonal) shows all positive correlations. Only the estimated accuracy of the acronym strategy correlated significantly with the actual accuracy of this strategy. The better the participants thought they had solved the problems with one strategy, the better they thought they had also solved the problems with the other strategy. Comparison of the correlation of the actual accuracy data showed that this was indeed an accurate perception of their performance (p = .074).
Evaluation time. The acronym and the rehearsal strategy did differ in conclusion evaluation time, $F(1, 18) = 7.55, p < .05$. Problems encoded with the acronym strategy were evaluated faster ($M = 2.8$ s) than those encoded with the rehearsal strategy ($M = 3.7$ s).

Choice phase

Performance on the three types of problems. In order to avoid violation of the sphericity assumption in repeated measures ANOVA containing repeated measures factors with more than two levels, a MANOVA was conducted to compare the accuracy and evaluation times of the unbiased and the biased problems. Planned comparisons with separate error terms for each comparison were used to compare specific means (Keppel, 1991).

The accuracy differed as a function of problem type, $F(2, 18) = 8.61, p < .01$. The unbiased problems ($M = 96\%$) and the problems biased towards acronyms ($M = 94\%$) were solved more correctly than the problems biased towards rehearsal ($M = 85\%$), respectively $F(1, 19) = 15.43, p < .001$ and $F(1, 19) = 11.85, p < .01$. The former two types of problems were solved equally accurately, $F < 1$.

Problem type also affected the evaluation time of the problems, $F(2, 18) = 4.14, p < .05$. The problems biased towards rehearsal were solved more slowly ($M = 3.0$ s) than the unbiased problems ($M = 2.7$ s), $F(1, 19) = 5.68, p < .05$, but equally as slowly as the problems biased towards acronyms ($M = 2.9$ s), $F < 1$. The evaluation time of the latter two types of problems differed marginally significantly from one another, $F(1, 19) = 3.75, p = .068$.

Strategy selection. The mean percentage of selection of the acronym strategy per problem type over all trials was calculated. The acronym strategy was frequently applied to the unbiased problems and to the problems biased towards acronyms; respectively in 83\% and 86\% of the cases. To the problems biased towards rehearsal, the acronym strategy was only applied in 10\% of the cases. In order to detect the underlying principles of strategy selection on the biased problems, logistic regression analyses were applied to the selection of the acronym strategy on the biased problems. The same predictors were used as in the analyses of Experiment 1 except for the predictor “problem set”, which was replaced by the within-subject predictor “problem feature”. The problem feature variable was coded 0 when the problems were biased towards rehearsal and 1 when the problems were biased towards acronyms. The only predictor that reached significance was problem feature, $\hat{\beta} = 1.08, \chi^2(1) = 18.41, p < .001, PRE = .56$, and this was also the only predictor that was included in the best-fitting model. Hence, the probability of selecting the acronym strategy was higher when the problems were composed of long premise terms (problem feature = 1), and thus were hard to encode with the rehearsal strategy, than when these problems
consisted of terms with similar starting letters (problem feature = 0) and were hard to encode with the acronym strategy.

Discussion

The results confirm the hypothesis that strategic processes guide premise encoding. When the problems were biased towards the acronym or the rehearsal strategy, in 86–90% of the cases, participants used the most appropriate strategy to encode these problems and they managed to do this in a very flexible manner, given that the problems were presented in a completely randomised order. The logistic regression analyses showed that only the problem features, determining how easily the strategies could be applied, directed the participants’ strategy choice.

Up until now, only simple connected (isotropic) linear reasoning problems had been presented. These problems can be readily encoded into a model. Once the relation between the first two premise terms is translated into a model, each new premise term can be added by simply positioning the term to the right of the premise term that was last added to the model. In addition, due to the constant sentence frame, reasoners even do not have to attend to the conjunction “before” that expresses the relation between the premise terms. Thus, model construction of these problems does not require complex operations. Now, it is possible that strategic decisions are only made when constructing a model does not require a large amount of processing resources and processing time. If complex model-construction operations must be accomplished, there may not be enough resources and/or time available to also pass through a strategic decision process. This was explored in the following two experiments.

EXPERIMENT 3

In this experiment, we complicated the model-construction process by using problems with a mixed sentence frame (“before” and “after”) and by presenting the premises in a semi-continuous order. Using a mixed sentence frame will, first of all, oblige reasoners to explicitly consider the relational term in each premise. Furthermore, following Clark (e.g., 1971), more resources will be needed to process the relation “after” than to process the relation “before”, since “after” is a marked term (e.g., Clark, 1969; Huttenlocher, 1968). It is also possible that, during the comprehension of the premises, the order of the premise terms is converted (e.g., “C after B”, “B before C”) to facilitate the positioning of the tokens in the model (e.g., Hunter, 1957; Huttenlocher, 1968). These conversions will also make the model-construction process more demanding.

Another aspect that will hamper model construction of the problems in this experiment is the semi-continuous presentation order of the premises. Previous research (e.g., Mynatt & Smith, 1977; Oakhill & Johnson-Laird, 1984; Potts, 1972) has shown that reasoning performance is better when the premises are
presented in a continuous order (e.g., AB, BC, CD) than when they are presented in a discontinuous (e.g., AB, CD, BC) or in a semi-continuous order (e.g., BC, CD, AB) (but see Ehrlich & Johnson-Laird, 1982). For the discontinuous order, this is due to the fact that up to a certain point (in the example: up to the second premise), no integrated representation can be constructed (e.g., Potts, 1972). For the semi-continuous order, on the other hand, an integrated representation can be constructed, but in order to add the information of the discontinuous premise (in the example: the last premise) in the model, this model must be mentally traversed to find the position where the new premise term must be attached. According to Sternberg (1980) this mental scanning requires an additional amount of processing time, thus complicating the construction process.

Method

Participants and design

A total of 24 participants were recruited by advertisements distributed in university buildings and paid 250 Belgian francs (6.20 Euro) for their participation in the experiments. None of them participated in the previous experiments. All participants applied the acronym and the rehearsal strategy (within-subject variable) to 10 problems each. The order of strategy application was a between-subjects variable.

Materials and procedure

A total of 52 temporal linear reasoning problems with a mixed sentence frame were constructed: 2 practice problems, $2 \times 10$ problems for the no-choice phase, and 30 problems for the choice phase: 10 unbiased problems, 10 problems biased towards acronyms, and 10 problems biased towards rehearsal. Of these 52 problems, 42 were constructed on the basis of the problems presented in Experiment 2, used in the same phases. Ten new problems, consisting of short premise terms, were created for application of the rehearsal strategy in the no-choice phase, such that both strategies could be applied to ten problems each.

Based on the results of a pilot study, it was decided to limit the number of premise terms to five. Hence, one premise term was omitted from the problems that had already been used in Experiment 3. Four different sentence frames were used to construct the problems (b = before, a = after): “C b D – E a D – B b C – B a A”, “B b C – C b D – E a D – B a A”, “D a C – C a B – A b B – D b E” and “C a B – A b B – D a C – D b E”. In each different problem set, the first two sentence frames occurred in three problems each and the other two sentence frames in two problems each. For each type of sentence frame, one problem was combined with a conclusion querying the relation between the second and the fourth term, and another problem with a conclusion querying the relation between the third and
the fourth term. The remaining two problems were either paired with a conclusion querying the relation between the first and the fourth term or between the second and the fifth term. In the conclusions, the conjunction “before” was used. The procedure of the experiment was the same as in Experiment 2 except that now the premises were presented for 10 seconds, since the results of the pilot study showed that 6 seconds was too short a time.

Results

No-choice phase

The accuracy of the strategies, the participants’ estimation of the accuracy of the strategies, and the mean evaluation time of the correctly solved problems were analysed by a 2 (strategy application order) × 2 (strategy type) ANOVA.

Accuracy. There were no main effects of strategy type or of strategy application order but their interaction was significant, $F(1, 22) = 11.85, p < .01$, reflecting a practice effect. When the acronym strategy was applied before the rehearsal strategy, more problems were solved correctly with the rehearsal strategy ($M = 97\%$) than with the acronym strategy ($M = 89\%$), whereas the opposite was true when the rehearsal strategy was applied first (respectively $M = 91\%$ and $M = 95\%$). Yet, the latter difference did not attain significance (Newman-Keuls post hoc comparisons).

Estimated accuracy. Analysis of the estimated accuracy data revealed a similar interaction effect of strategy type and strategy application order, $F(1, 22) = 5.31, p < .05$. When the acronym strategy was applied first, the estimation of the accuracy of the acronym strategy ($M = 63\%$) was lower than when the rehearsal strategy was applied first ($M = 75\%$). The accuracy of the rehearsal strategy was estimated equally high, independent of the order of application (first $M = 68\%$, second $M = 70\%$) (Newman-Keuls post hoc comparisons).

A correlation analysis on the accuracy and the estimated accuracy of both strategies (Table 5) shows only positive correlations. The estimated accuracy of the acronym and of the rehearsal strategy correlated significantly with the actual accuracy of the respective strategies. Similar to the previous experiment, the estimations of the accuracy of both strategies correlated significantly.

Evaluation time. There was again only an interaction between strategy type and strategy application order, $F(1, 22) = 10.36, p < .01$. The problems were solved faster with the rehearsal ($M = 4.3$ s) than with the acronym strategy ($M = 5.2$ s), when the acronym strategy was applied first, but faster with the acronym strategy ($M = 4.2$ s) than with the rehearsal strategy ($M = 5.0$ s) when the rehearsal strategy was applied first (Newman-Keuls post hoc comparisons).
Choice phase

Performance on the biased and unbiased problems. The accuracy and evaluation time of the biased and the unbiased problems were analysed by MANOVAs. There were no differences in accuracy, $F(2, 22) = 1.25, p = .305$, nor in evaluation time, $F(2, 22) = 2.10, p = .146$, between the unbiased problems ($M = 95\%; M = 3.7 \text{ s}$), the problems biased towards acronyms ($M = 93\%; M = 4.0 \text{ s}$), and the problems biased towards rehearsal ($M = 91\%; M = 4.3 \text{ s}$).

Strategy selection. The mean percentage of acronym strategy selection per problem type over all trials was calculated. The acronym strategy was applied to 52% of the unbiased problems, to 74% of the problems biased towards acronyms, and to 14% of the problems biased towards the rehearsal strategy.

The selection of the acronym strategy on the biased problems was again submitted to logistic regression analyses using the same predictors as in Experiment 2. Problem feature was once again the most important predictor, $\beta = .88, \chi^2(1) = 17.82, p < .001$. $PRE = .42$, followed by the difference in estimated accuracy, $\beta = .29, \chi^2(1) = 3.04, p = .081$, $PRE = .06$ and the strategy estimated to be the most accurate one $\beta = .28, \chi^2(1) = 4.07, p < .05$. The estimated accuracy of the acronym strategy, $\beta = .22, \chi^2(1) = 3.24, p = .072$, $PRE = .03$, only marginally significantly predicted strategy choice.

The best-fitting model, $PRE = .52$, included two predictors: problem feature, $\beta = 1.02, \chi^2(1) = 17.82, p < .001$, and the strategy estimated to be the most accurate one, $\beta = .52, \chi^2(1) = 4.07, p < .05$. This model shows that there was a higher probability of selecting the acronym strategy when the problems consisted of long premise terms (problem feature = 1) than when the problems consisted of terms with similar starting letters (problem feature = 0). The probability of applying the acronym strategy was also higher when the acronym strategy was estimated to be the most accurate strategy.

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Rehearsal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Estimated accuracy</td>
</tr>
<tr>
<td>Acronyms</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Estimated accuracy</td>
<td>—</td>
</tr>
<tr>
<td>Rehearsal</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Estimated accuracy</td>
<td>—</td>
</tr>
</tbody>
</table>

"** $p < .001$, * $p < .05$
Discussion

The aim of the present experiment was to investigate whether reasoners still manage to strategically encode reasoning problems that require a more complex model construction. First, although the model-construction process of these problems was presumed to be quite complex due to the mixed sentence frame and the semi-continuous presentation of the premises, the problems were solved highly accurately and relatively quickly. One reason for this high level of performance may be that the participants of this study might have been more motivated than the participants in the previous studies, because they were paid for their participation. It is also possible that the longer premise-processing time (10 s) allowed the participants to construct a more stable model on the basis of which more accurate inferences could be made. The complexity of the model-construction process, however, does not necessarily have to be reflected in the accuracy or evaluation time of the problems. Potts and Scholz (1975), for instance, established that conclusion evaluation times were unaffected by the markedness of the relation used in the premises. Once the premise information was represented, the way in which this information was formulated no longer affected reasoning performance. However, the complexity of the model-construction process will influence the time needed to process the premises. Yet, because this time is fixed in the current experiments, this information is not available.

Second, the strategy selection data revealed that reasoners did indeed select the most efficient strategy to encode the problems. Problem feature again determined strategy choice, but now the participants’ estimation of the most accurate strategy was also included in the best-fitting model. Similar to Experiment 1, only a very elementary appraisal of the strategies, namely the fact that one strategy is considered to be better than the other one, guides strategy selection. Note also that, although the order in which the strategies were applied in the no-choice phase influenced the perceived accuracy of the acronym strategy, strategy application order did not affect strategy selection.

Although problem feature is the most important predictor of strategy application in the current experiment, the average percentage of acronym application to the problems biased towards acronyms \( (M = 74\%) \) was significantly lower than the corresponding percentage in the previous experiment \( (M = 86\%) \), \( \chi^2(1) = 8.88, p < .01 \). However, this was not the case for the application of the rehearsal strategy to the problems biased towards rehearsal (respectively \( M = 86\% \) vs \( M = 90\%) \), \( \chi^2(1) = 2.18, p = .14 \). Closer inspection of the data showed that the difference in acronym selection on the problems biased towards acronyms may be due to four participants who consistently (at least 29 times of 30) applied the rehearsal strategy to all problems in the choice phase. When these participants, as well as those who consistently applied acronyms to all problems \( (n = 3) \), and the single participant who consistently applied the
rehearsal strategy in the previous experiment, were removed from the analyses, the difference in acronym application to the problems biased towards acronyms between both experiments was no longer significant (respectively $M = 86\%$ vs $M = 90\%$), $\chi^2(1) = 1.50$, $p = .221$. Hence, the difference in acronym application to the problems biased towards acronyms between this experiment and the previous experiment seems to be due to the fact that some participants preferred to apply one strategy, the one they considered to be the most accurate one, consistently to all problems. Presenting more complex problems thus seems to reveal individual differences in adaptivity.

EXPERIMENT 4

In the previous experiment, the complexity of the problems was increased by the way in which the relation between the premise terms was described. In the current experiment we varied the complexity by manipulating the nature of this relation. More specifically, we presented one-model problems, describing a determinate relation between the premise terms, and two-model problems describing an indeterminate relation. Recent research on relational reasoning has shown that model construction of multiple-model problems remains of a more complex nature than model construction of one-model problems, even when the multiple possibilities are represented in only one model (Schaeken & Van der Henst, 2002; Vandierendonck, De Vooght, Desimpelaere, & Dierckx, 2000). According to these studies, reasoners seem either to represent all possible relations of a multiple-model problem immediately in the initial model (isomeric model; Schaeken & Van der Henst, 2002) or to represent only one possible relation, elaborated with a marker indicating that, from that point on, the relation is indeterminate (annotated mental model; Vandierendonck et al., 2000). These operations require a supplementary amount of resources and processing time, complicating the model-construction process of two-model problems. Also conclusion evaluation is influenced by the nature of the relation described in the problems. However, Vandierendonck et al. (2000) have shown that this is only the case for conclusions querying an indeterminate relation (indeterminate conclusions) of the multiple-model problems. In order to evaluate these conclusions, reasoners have to flesh out the marker into a forked structure that makes explicit the multiple relations between the premise terms. This process requires an extra amount of resources, which are not required to evaluate a conclusion of a one-model problem or to evaluate a conclusion querying a determinate relation (determinate conclusion) of a multiple-model problem.

In the present experiment, we presented one-model problems, two-model problems with a determinate conclusion, and two-model problems with an indeterminate conclusion. Because the acronym and the rehearsal strategy have only been applied to one-model problems (see also Mynatt & Smith, 1977), it was not clear whether both strategies were also suited to encode two-model
problems into an isomeric (Schaeken & Van der Henst, 2002) or an annotated model (Vandierendonck et al., 2000). Therefore, we also included a control condition in which reasoners had to solve the same problems as in the experimental condition but without strategy instruction.

Method

Participants and design

A total of 61 first-year students of the Faculty of Psychology and Educational Sciences at Ghent University participated for course requirements and partial credit. None of them had participated in the previous experiments. They were randomly assigned to the control condition (n = 30) or to the experimental condition (n = 31). The participants in the experimental condition had either to apply the acronym strategy before the rehearsal strategy or the rehearsal strategy before the acronym strategy. A similar between-subjects variable was created in the control condition. Participants had either to solve the problems presented for application of the acronym strategy in the experimental condition (problem set acronyms) before the problems presented for application of the rehearsal strategy (problem set rehearsal) or in the opposite order. In order to use the same terminology in both conditions, now the terms “problem set” and “strategy application order” will be used instead of “strategy” and “strategy application order”. The design is thus a 2 (condition: experimental vs control) × 2 (problem set order) × 2 (problem set: problem set acronyms vs problem set rehearsal) × 2 (number of models: one-model problems vs two-model problems) design with the first two factors between-subjects and the last two factors within-subject.

Materials

Six one- and six two-model problems were created for each problem set. Hence, there were 12 problems for application of the acronym strategy and 12 problems for application of the rehearsal strategy in the no-choice phase, and 12 unbiased problems, 12 problems biased towards acronyms, and 12 problems biased towards rehearsal for the choice phase. The same problems were used as in the previous experiment with a few changes. First, we reused a constant sentence frame (“before”) to describe the relation and for some of the two-model problems we reordered the terms such that two Dutch-sounding pseudo-words could be formed. Furthermore, in half of the problems a determinate relation (one-model problems) and in the other half an indeterminate relation (two-model problems) was described. The last premise indicated whether the relation was determinate or indeterminate. In each problem set, two new problems were created to obtain 12 problems per set.

Two types of conclusions were paired with the problems: determinate conclusions and indeterminate conclusions. The determinate conclusions were
identical for the one- and the two-model problems. These conclusions queried the
determinate relation between either the second and the fourth premise term (BD
and DB), the third and the fifth term (CE), or the third and the fourth premise
term (DC). The remaining two two-model problems were paired with an
indeterminate conclusion querying the relation between the fourth and the fifth
premise term (DE and ED). The remaining two one-model problems were paired
with the conclusion ED or CA. These two problems were used as filler problems,
to maintain an equal number of one- and two-model problems, and were not
included in the analyses.

Procedure

The procedure of the experimental condition was identical to the procedure of
Experiments 2 and 3, with two minor changes: the content of the general
instructions and the fact that three answer alternatives were provided. In the
present general instructions, an example of a two-model problem, paired with an
indeterminate conclusion, was presented. Participants were told that there were
easy and more difficult problems, and it was explained that the relation in the
conclusion of the difficult (two-model) problem in the example was
indeterminate because both orders of the terms could be possible. Furthermore,
they were instructed to press the middle mouse button, meaning “no
unambiguous answer”\(^1\) when they believed they had insufficient information to
evaluate a conclusion. Correct conclusions still had to be indicated by pressing
the left mouse button and incorrect conclusions by pressing the right mouse
button. Then the specific instructions followed, either explaining the acronym
strategy or the rehearsal strategy by means of a one-model problem. After that it
was briefly mentioned not to forget that there were also more difficult reasoning
problems but it was left aside how these problems could be encoded with the
strategies. In the control condition, the same general instructions were presented
but no specific instructions were given. Unlike in the experimental condition,
participants were allowed to work in silence. In all other respects, the procedure
of both the experimental and the control conditions were identical to the
procedure used in the previous experiment.

Results

No-choice phase

For each condition, the mean accuracy and the mean evaluation time for the
one-model and the two-model problems with a determinate conclusion (averaged
over four problems) were calculated, separately for the two problem sets. These
means were submitted to a 2 (condition) × 2 (problem set order) × 2 (problem set)

\(^1\)In the Dutch language this was expressed as “geen eenduidig antwoord”, which does not
include a double negative.
MODEL CONSTRUCTION STRATEGIES

× 2 (number of models) ANOVA with the first two variables between-subjects and the last two variables within-subject. We also calculated the mean accuracy and the mean evaluation time for the two-model problems having an indeterminate conclusion (average over two problems). These means were submitted to a 2 (condition) × 2 (problem set order) × 2 (problem set) ANOVA.

**Accuracy.** The problems with a determinate conclusion were solved more correctly in the experimental than in the control condition (M = 93% vs M = 87%), F(1, 57) = 4.64, p < .05, and one-model problems were solved more accurately than two-model problems (M = 94% vs M = 85%), F(1, 57) = 25.31, p < .001. There was also an interaction of problem set with problem set order, F(1, 57) = 4.89, p < .05. The problems used for application of the acronym strategy were solved less accurately when these were solved before (M = 85%) than after (M = 91%) the problems of the rehearsal strategy. Presentation order did not affect the accuracy of the problems used for application of the rehearsal strategy (M = 90% vs M = 93%) (Newman-Keuls post hoc comparisons).

For the problems with an indeterminate conclusion there was only a main effect of condition, F(1, 57) = 10.94, p < .01. Problems with an indeterminate conclusion were solved more accurately in the control than in the experimental condition (M = 62% vs M = 31%).

**Evaluation time.** The one-model problems with a determinate conclusion (M = 4.7 s) were solved faster than the two-model problems (M = 5.1 s), F(1, 57) = 8.70, p < .01. Analysis of the problems with an indeterminate conclusion only showed an interaction between problem set and problem set order, F(1, 57) = 5.27, p < .05. The problems presented for application of the acronym strategy were evaluated more slowly when they were presented before (M = 5.5 s) than after (M = 4.4 s) the problems presented for application of the rehearsal strategy. However, the evaluation time of problems used for application of the rehearsal strategy did not vary as a function of the order in which the problem sets were presented (M = 4.9 s vs M = 4.8 s) (Newman-Keuls post hoc comparisons).

**Estimated accuracy.** The estimated accuracy of the acronym and the rehearsal strategy was equally high (respectively M = 84% and M = 86%) and did not vary as a function of the order in which the strategies were applied, F(s) < 1.

**Choice phase**

**Performance on the unbiased and biased problems.** The mean accuracy and the mean evaluation time of the correctly solved unbiased and biased problems (problem type) are presented in Table 6. Separate MANOVAs were conducted for the problems with a determinate conclusion and for the problems with an indeterminate conclusion, with condition as a between-subjects variable. The
results were similar to the results of the no-choice phase. Therefore, only the findings involving problem type are reported.

The accuracy data of the problems with a determinate conclusion showed a three-way interaction between condition, problem type, and the number of models, $F(2, 58) = 4.03, p < .05$. The unbiased problems were solved more accurately than the biased problems, except for the two-model problems in the control condition. Also, the evaluation times of these problems showed an effect of problem type, $F(2, 58) = 6.72, p < .01$, but now the conclusions of the unbiased problems and of the problems biased towards acronyms were evaluated faster than the conclusions of the problems biased towards rehearsal, respectively $F(1, 59) = 13.42, p < .001$ and $F(1, 59) = 7.39, p < .01$. The first two types of problems did not differ in evaluation time, $F < 1$.

For the problems with an indeterminate conclusion the evaluation time also differed as a function of the type of problem presented, $F(2, 58) = 4.83, p < .05$. Only the conclusions of the unbiased problems ($M = 3.8$ s) were evaluated faster than the conclusions of the problems biased towards rehearsal ($M = 4.7$ s), respectively $F(1, 59) = 9.36, p < .01$. The difference in evaluation time between the unbiased problems and the problems biased towards acronyms ($M = 4.2$ s) was not significant, $F < 1$, nor was the difference between the biased problems, $F(1, 59) = 2.96, p = .090$.

### TABLE 6
Mean accuracy (ACC) and mean evaluation time (ET; seconds) per problem type in the experimental and control conditions in the choice phase of Experiment 5

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>ACC</th>
<th>ET</th>
<th>ACC</th>
<th>ET</th>
<th>ACC</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unbiased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-model problems</td>
<td>.98</td>
<td>3.8</td>
<td>.92</td>
<td>4.0</td>
<td>.94</td>
<td>4.6</td>
</tr>
<tr>
<td>Two-model problems with determinate conclusions</td>
<td>.92</td>
<td>4.3</td>
<td>.89</td>
<td>4.3</td>
<td>.84</td>
<td>4.6</td>
</tr>
<tr>
<td>Two-model problems with indeterminate conclusions</td>
<td>.42</td>
<td>3.9</td>
<td>.40</td>
<td>4.4</td>
<td>.36</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>Biased towards acronyms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-model problems</td>
<td>.98</td>
<td>3.8</td>
<td>.90</td>
<td>4.2</td>
<td>.87</td>
<td>4.2</td>
</tr>
<tr>
<td>Two-model problems with determinate conclusions</td>
<td>.83</td>
<td>4.7</td>
<td>.85</td>
<td>4.7</td>
<td>.87</td>
<td>4.8</td>
</tr>
<tr>
<td>Two-model problems with indeterminate conclusions</td>
<td>.75</td>
<td>3.8</td>
<td>.65</td>
<td>4.0</td>
<td>.65</td>
<td>4.9</td>
</tr>
</tbody>
</table>
Strategy selection. Based on the verbalisations made by the participants, their strategy use on the three types of problems was detected. Furthermore, it was also noted how the participants applied the acronym and the rehearsal strategy to the two-model problems. For 58% of the participants, there was no difference between the verbalisations they made while solving one-model problems and while solving two-model problems. In fact, 10 of these 18 participants consistently responded to the problems with indeterminate conclusions as if the order between the terms was an ABCDE order. On average 77% of the responses of the participants in this group were in accordance with this ordering. The remaining 42% of the participants could be partitioned as follows: 13% constructed two separate models, 13% constructed an isomeric or annotated model (revealed by mentioning an “or” between the last two premise terms or by leaving some time between the encoding of the first three premise terms and the last two terms), and 16% constructed a model of the first four premise terms and added a mini-string (see Vandierendonck, Dierckx, & De Vooght, 2002) comprising the premise term that is repeated and the last term. Only in the last group did two participants consistently respond to the indeterminate conclusions as if the order was an ABCDE order. On average 53% of the responses of these participants was in accordance with this ordering. In the other two groups, on average only 20% of the responses accorded with an ABCDE ordering of the terms.

Because the level of performance on the problems with indeterminate conclusions was so low and the verbalisations of 58% of the participants did not reveal a difference in encoding between the one- and the two-model problems, we suspected that some participants simply did not notice the indeterminacy of the two-model problems. Hence, two groups of participants were formed, based on the criterion that more than three of the six indeterminate conclusions in the choice phase should have been solved correctly: a group of participants who were believed to have noticed the indeterminacy of the two-model problems and a group who were believed not to. Table 7 shows the percentage of acronym selection as a function of the number of models and whether or not the participants noticed the indeterminacy of the two-model problems. As can be seen in the table, participants who noticed the indeterminacy of the two-model problems applied the acronym strategy less frequently to all six types of problems.

Because we are only interested in the strategy application of the participants for whom the two-model problems were more complex than the one-model problems, we only conducted logistic regression analyses on the acronym selection data of the biased problems of the participants who noticed the indeterminacy. Two types of logistic regression analyses were conducted: one to
explore the determinants of acronym application to the one-model problems and
the other to the two-model problems. The same predictors were used as in the
previous two experiments, but now specified for the one- and the two-model
problems. The mean accuracy and the mean evaluation time of the one-model
problems and of the two-model problems with a determinate conclusion in the
no-choice phase were used as predictors of acronym application to respectively
the one-model problems and the two-model problems in the choice phase. Also
the predictors derived from the accuracy scores and the evaluation times
difference scores and most accurate or fastest strategy) were specified for the
one-model and two-model problems. Only the predictors based on the
estimations of the participants were the same for the one- and for the two-model
problems.

The individual logistic regressions for the one- and the two-model problems
are displayed in Table 8.\(^3\) The best-fitting model for the one-model problems
included only the predictor “problem feature”. There was thus again a higher
probability of selecting acronyms when the one-model problems were biased
towards acronyms than when these problems were biased towards rehearsal. The
best-fitting model for the two-model problems included two predictors, \(PRE = .58\). The most influential predictor was again problem feature, \(\beta = 1.54, \chi^2(1) = 8.63, p < .01\). The other predictor was the actual accuracy of the acronym strategy
for the two-model problems, \(\beta = 1.06, \chi^2(1) = 5.20, p < .05\). So, for the two-

\(^3\)One participant was excluded from the analyses because his/her data were completely
different from the data of the other participants.
model problems also there was a higher probability of selecting acronyms when the problems were biased towards acronyms than when the problems were biased towards rehearsal. Furthermore, the higher the actual accuracy of the problems solved with the acronym strategy, the higher the probability that this strategy was selected in the choice phase.

Discussion

The present experiment investigated whether strategic processes also play a role in the encoding of two-model problems that have a more complex model construction than one-model problems. From the results of the no-choice phase, it appeared that the acronym and the rehearsal strategy are good strategies for encoding the one-model problems but are not very well suited for encoding multiple-model problems. Indeed, it was found that reasoners were able to solve the one-model problems more accurately when they were taught the acronym or the rehearsal strategy than when no strategy was taught, as was the case in the control condition. However, the opposite pattern was found for the indeterminate conclusions of the two-model problems. These conclusions can only be evaluated correctly when reasoners represent the indeterminacy of the relation in the mental model, whereas the determinate conclusions of these problems can also be solved correctly if only one instead of two possible relations between the premise terms is taken into account. By applying the acronym or the rehearsal strategy,
reasoners seemed to be less likely to take into account the indeterminacy of the problems during conclusion evaluation. This may be due to the fact that a constant sentence frame is used to describe the relation in which every last premise term is also the first term of the subsequent premise. Because of this, it is unnecessary to read the entire premises and two short-cut versions of the strategies may be used instead. One short-cut version consists of scanning only the first terms of the first three premises and the entire last premise to encode the fifth term. The other version consists of the complementary situation in which the first premise is scanned in its entirety and then only the last terms of the following premises. When this latter short-cut version of the strategies is used, reasoners will not notice the indeterminacy of the two-model problems because this is realised by repeating the first term of the third premise as the first term in the fourth premise. So, scanning only the last premise terms will not disclose that the problem is a two-model problem. Scanning only the first terms, on the contrary, will reveal the indeterminacy because then the repetition of the term is noticed.

An extra analysis, comparing the accuracy of these problems between the group of reasoners who recognised the indeterminacy and the control group, did indeed reveal that, when the indeterminacy was recognised, performance on this type of problem no longer differed between the groups, $F(1, 42) = 1.40, p = .247$. If anything, there was even a tendency for the experimental group ($M = 81\%$) to solve more problems correctly than the control group ($M = 68\%$).

The strategy choice data of the reasoners who noticed the indeterminacy of the two-model problems again revealed that the features of the problems directed strategy application. However, the percentage of acronym application to the two-model problems biased towards acronyms only amounted to 54%. This percentage was significantly lower than the percentage for the problems in the previous two experiments (Exp. 2: $M = 86\%$; Exp. 3: $M = 74\%$), respectively $\chi^2(1) = 31.05, p < .001$ and $\chi^2(1) = 11.67, p < .001$. To check whether this lower percentage was also due to individual differences in adaptivity, the percentage of acronym selection on problems biased towards acronyms was recalculated while excluding those participants who consistently applied the rehearsal ($n = 3$) or the acronym strategy ($n = 1$) to all problems. Yet even without these participants, the difference in acronym selection between the present experiment ($M = 73\%$) and Experiment 2 ($M = 90\%$), $\chi^2(1) = 9.36, p < .01$, as well as between the present experiment and Experiment 3 ($M = 86\%$), $\chi^2(1) = 4.46, p < .01$, remained. This seems to suggest that also within individuals, strategy switches are less adaptive when more complex two-model problems are presented. However, this rigidity was only limited to the problems biased towards acronyms. The percentage of acronym application to the two-model problems biased towards rehearsal was significantly lower than to the problems in the previous experiments (Exp. 4: $M = 4\%$; Exp. 2: $M = 15\%$; Exp. 3: $M = 10\%$), respectively $\chi^2(1) = 3.78, p = .052$ and $\chi^2(1) = 9.02, p < .01$. Increasing the complexity of the problems thus seems to prompt reasoners to behave less adaptively, thereby increasing the use of the rehearsal strategy.
In sum, the data of the present experiment reveal that reasoners are still able to switch strategies in response to changing problem features when more complex two-model problems are presented. However, the complexity of the problems seems to bring about a less adaptive application of the strategies, particularly increasing the use of the rehearsal strategy.

**GENERAL DISCUSSION**

The results of all four experiments corroborate the hypothesis that model construction of linear reasoning problems is open to strategic decisions. In general, reasoners were found to select the most suitable strategy to encode the problems, and this strategy selection occurred in a very flexible manner adaptive to changing problem features. Although problem feature was still the main predictor, the adaptiveness of strategy application decreased when model construction became harder (Experiments 3 & 4). This observation is not unexpected because people have only limited processing resources available, which they have to distribute among the strategic-decision and the model-construction process within a limited space of time. In order to ascertain an adaptive strategy application, the strategy decision process should consist of a scanning and processing of the problem features followed by the selection of the strategy that is most suited to encode these problems with the features encountered (e.g., Lemaire & Siegler, 1995; Siegler & Lemaire, 1997). To accomplish these operations, a certain amount of processing resources and processing time is required, which may even be increased when participants have to switch from one strategy—applied to the previous problem—to the other (switch cost, e.g., Allport, Styles, & Hseih, 1994; Rogers & Monsell, 1995). This amount of resources and/or time, together with the resources and the time required for the more complex model construction, may exceed one’s capacity and may prompt reasoners to simplify the strategic decision process. As there are large individual differences in working memory capacity and processing speed, the appearance of individual difference in adaptivity (see Experiments 3 & 4) is indeed to be expected.

Increasing the complexity of the problems resulted in a more frequent application of the rehearsal strategy. Presumably this is because the rehearsal strategy is to some extent suited to encoding all problems, even the problems comprising long premise terms (see Experiment 1). Selecting the rehearsal strategy is thus a very safe choice that does not have to be preceded by an evaluation of the problem features, because there are no features that substantially hamper the encoding of the problems with the rehearsal strategy. Even more, omitting this evaluation will leave additional processing time that may be used to construct a more stable representation. Hence, when solving problems requires a vast amount of resources and processing time, reasoners sometimes seem to prefer a sufficiently efficient strategy to a strategy that may be more suited to encoding the problems.
The augmented complexity of the model-construction phase does not seem to be the only factor that brings about a less adaptive strategy application. In Experiment 1, it was found that easy problems (i.e., problems with short premise terms starting with different letters) were solved more accurately with the acronym than with the rehearsal strategy. Yet, as observed in the same experiment, reasoners who applied the acronym strategy to 10 hard problems and the rehearsal strategy to 10 easy problems in the no-choice phase, selected the rehearsal strategy for application to the easy problems in the choice phase. Strikingly, although instructed to solve the problems as accurately as possible and despite the fact that reasoners knew that the “nature” of the first letters was a relevant problem feature, as revealed by their comments in the no-choice phase, 14 out of 15 participants preferred the less efficient rehearsal strategy to encode the easy problems. This finding is related to the *Einstellung* effect, which was first investigated by Luchins (1942). He observed that after having solved a sequence of training problems by means of the same method, people kept on applying this method to a critical problem that could be solved by a much simpler method. However, participants who did not receive the training problems solved the critical problem immediately with the simpler method. Similarly, here applying the acronym strategy to 10 hard problems and the rehearsal strategy to 10 easy problems might have created such a “mental set”, blinding the participants to the possibility that the acronym strategy might be more appropriate than the rehearsal strategy to encode the problems in the choice phase. Which factor(s) is (are) responsible for this non-adaptive strategy application, however, still needs to be determined and may be important since research on strategy selection mostly stresses the adaptive nature of strategy application (e.g., Geary & Wiley, 1991; Kerkman & Siegler, 1993; Siegler & Lemaire, 1997).

The present study suggests that reasoners have some cognitive control of which model they construct from the premises and of how many resources they invest in its construction. To a certain extent, this idea is already present in the work of Evans (2000b) and Klauer, Musch, and Naumer (2000) on the belief bias effect in syllogistic reasoning. This effect refers to the tendency to accept believable conclusions and to reject unbelievable ones irrespective of their validity. These researchers suggest that the effect is due to reasoners trying to construct a model that sustains the conclusion when this conclusion is believable, and trying to construct a model in which the conclusion does not hold when the conclusion is unbelievable. Reasoners thus seem to be able to tailor their model construction to the believability of the conclusion, such that a conclusion can be drawn that does not endanger their belief system (Evans, 2000b). Within the context of the belief bias effect, Cherubini, Garnham, Oakhill and Morley (1998) also provided some tentative evidence showing that reasoners may be able to strategically control model construction. According to them, belief bias effects arise because reasoners retrieve a model from their long-term memory that
represents the relation in the conclusion. Starting from this long-term memory model, reasoners try to add all the information from the premises in this model. If this is possible, the conclusion is judged to be valid. However, when the premise information cannot be incorporated in the model, Cherubini et al. found that the belief bias effect disappeared. They argued that this is due to reasoners discarding the long-term memory model, to start constructing a model from the premise information solely. In other words, reasoners seem to be able to suppress their beliefs during model construction, if they realise that the information to be processed conflicted with their beliefs. However, more research is still needed to test the generality of this effect and to further reveal strategic processes in model construction.

REFERENCES


