

Reasoning about the relations between relations

Geoffrey P. Goodwin and P. N. Johnson-Laird

Princeton University, Princeton, NJ, USA

Relations can hold between relations, as in assertions such as: Cordelia loves Lear more than Goneril does. Naïve reasoners can make inferences that depend on these higher order relations, which are vital for science and mathematics, but no existing theory explains such inferences. The present paper presents a theory based on mental models of the situations under description, and it reports four experiments corroborating the theory. Experiment 1a showed that the difficulty of such inferences from two premises depends on the integration of the information from the premises into a single model. The same result held in Experiment 1b, even when individuals were not permitted to make written workings. Experiment 2 required the participants to think aloud, and their protocols revealed that they developed three main strategies. Experiment 3 biased the development of these strategies, showing that individuals assemble them “bottom up” from various tactical steps.

Reasoning about relations is widespread. Many inferences in daily life depend on temporal, spatial, and other sorts of relation. Relations differ in the number of arguments that they take. A relation such as *taller than* takes two arguments—for example, Pat is taller than Viv. A relation such as *between* takes three arguments—for example, Pat is between Viv and Joe. And a relation such as *swap* can take four arguments—for example, Pat swapped a bicycle for a scooter with Viv. The American logician C.S. Peirce analysed the logic of relations. He showed that any relation can be reduced to a set of relations in which none takes more than three arguments, and he developed diagrams for representing relations that are remarkably similar to modern semantic networks (see Hartshorne, Weiss, & Burks, 1931–1958, e.g., Vol. 3, paragraph 469).

The number of arguments in a relation, according to Halford and his colleagues, is likely to affect the difficulty of understanding the relation and in using it to make inferences (see, e.g., Halford, Wilson, & Phillips, 1998). These authors argued that the difficulty of a task depends on its *effective* relational complexity—that is, on the most complex relation that has to be processed in carrying out the task, where complexity is solely a matter of the number of arguments. They reviewed developmental research corroborating this claim. Other studies have shown that relational complexity increases the difficulty of certain sorts of deductive reasoning, including syllogistic reasoning and metalogical reasoning (the “knights and knaves” task, see Birney & Halford, 2002).

Psychologists have focused on binary relations, and they have proposed a variety of theories to

Correspondence should be addressed to Geoffrey P. Goodwin, Department of Psychology, Princeton University, Princeton, NJ 08544, USA. Email: ggoodwin@princeton.edu

This research was supported by a grant from the National Science Foundation (BCS-0076287) to study strategies in reasoning. For their helpful advice, we thank Ruth Byrne, Sam Glucksberg, Graeme Halford, Uri Hasson, Cathy Haught, Louis Lee, Tom Ormerod, Sanna Reynolds, Mike Oaksford, and two anonymous reviewers.

explain inferences with them (e.g., Clark, 1969; Huttenlocher, 1968). Recent accounts suggest that naïve reasoners rely on mental models of the various possibilities to make such inferences (see, e.g., Byrne & Johnson-Laird, 1989; Carreiras & Santamaría, 1997; Schaeken, Johnson-Laird, & d'Ydewalle, 1996). None of the theories, however, has considered a special sort of complexity that can occur in relations.

There can be relations between relations. They occur in a variety of domains. In daily life, people assert that a relation holds to a greater degree between one pair of individuals than another. The tragedy of King Lear revolves around such a familiar pattern. Which of Lear's daughters loves him the most—does Cordelia love him more than Goneril loves him or more than Regan loves him? King Solomon's reputation for wisdom is based in part on his ability to determine which of two women loved a child more. Likewise, an analogy usually depends on the similarity (a relation) between one relation and another, and questions about analogies are common in intelligence tests—for example, steel is to rust as bread is to what? Psychologists also encounter relations between relations when they consider an interaction between variables (see, e.g., Halford, Baker, McCredden, & Bain, 2005). An ability to deal with these higher order relations is crucial to a number of complex spatio-temporal skills such as air traffic control (e.g., Loft, Hill, Neal, Humphreys, & Yeo, 2004). And a simple relation between relations can have vast implications in mathematics. For instance, divide a line into two parts so that the ratio of the longer segment to the shorter segment is the same as the ratio of the line as a whole to the longer segment. This ratio, which is a relation between relations, is one of the most famous irrational numbers: ϕ (the golden ratio).

How do people reason about relations between relations? Consider, for example, the result of a recent study of the lead-up to the 2004 US Presidential election (Kull, 2004):

Bush supporters were less knowledgeable about his policies than were Kerry supporters about Kerry's policies.

Let us assume the further premise:

Voters are more disappointed the greater the discrepancy between their beliefs about a candidate's policies and the candidate's actual policies.

You can infer that Bush's supporters would eventually be more disappointed if he were to be elected than Kerry's supporters would be disappointed if he were to be elected. No one knows how you make such inferences. Hence, our aim in the present paper is to take a first step towards answering the question. We needed to study inferences that necessarily depend on relations between relations, and we devised two sorts of problem to examine the underlying mental processes. The problems were based on pairs of premises that describe as succinctly as possible relations between relations, for example:

Art is taller than Bill to a greater extent than Cal is taller than Dale.

Bill is taller than Dale to a greater extent than Cal is taller than Dale.

What is the order of the four individuals in terms of their height?

To ensure that our participants had to reason, we used a task in which they put the four individuals into order according, in this case, to their height. Hence, these problems depend on pairs of assertions that both state relations between relations. Such problems are artificial and are unlikely to occur in everyday life. Nevertheless, they allow us to study reasoning about relations between relations.

We can symbolize an assertion, such as the first premise in the preceding example, in the following way:

$$(A > B) > (C > D)$$

where we use the inequality between the two clauses to assert that the difference between the two items in the first clause is greater than the difference between the two items in the second clause—that is $(A - B) > (C - D)$. Of course, the subtractions in this algebraic

formulation fail to convey that A is bigger than B and that C is bigger than D, and so the natural language expression conveys two relations and the relation between them.

We symbolize the two premises in the problem as follows:

$$(A > B) > (C > D)$$

$$(B > D) > (C > D)$$

The reader is invited to put the four individuals in order according to their height. The solution is straightforward. For example, you can construct a model from the first premise representing that Art is taller than Bill:

Art

Bill

where the vertical axis represents relative height. You can also construct a model representing that Cal is taller than Dale:

Cal

Dale

You cannot combine the two models into a single integrated model of the relative heights of the four individuals, but you do know that Art is taller than Bill to a greater extent than Cal is taller than Dale. Hence, you have two pieces of information about the same situation, and so you have two parts of the same model, but you cannot yet fit them together:

| | | |
|------|--|------|
| Art | | |
| | | Cal |
| | | |
| | | Dale |
| Bill | | |

The vertical line demarcates the two parts of the same model, which shows the relation between the two relations—that is, Art is taller than Bill to a greater extent than Cal is taller than Dale. The second premise allows you to construct a separate ordering of three individuals. It tells you that Bill is taller than Dale to a greater

extent than Cal is, and so the order of the four individuals is:

Art

Bill

Cal

Dale

If individuals use models to solve these problems, then there is a way to solve this problem that should be easier. You start with the second premise, which asserts that both Bill and Cal are taller than Dale, and further that Bill is taller than Dale by a greater amount than Cal is. Hence, Bill is taller than Cal, and so it follows:

Bill

Cal

Dale

This model can then be integrated with the information in the first premise. Its first clause stipulates that Art is taller than Bill, and so now you can construct a model of the complete rank order of heights:

Art

Bill

Cal

Dale

The first premise, of course, contains further information, but it is not needed for the present problem. All that is needed is a relation from a single clause of the first premise (here, the first clause), and so henceforth we refer to this sort of problem as a *single-clause* problem.

The second sort of problem is exemplified by the following premises:

Art is taller than Bill to a greater extent than Cal is taller than Dale.

$$(A > B) > (C > D)$$

Cal is taller than Dale to a greater extent than Bill is taller than Dale.

$$(C > D) > (B > D)$$

The second premise again yields an ordering of three individuals when the relation between the two clauses is considered:

Cal

Bill

Dale

A single relation from the first premise no longer suffices to rank order the four individuals. You need to take into account both relations in the first premise and the relation between them—that is, you need to take into account all the information in the model of the first premise (and the model of the second premise). Art must be the tallest individual to ensure that the difference between Art and Bill is greater than the difference between Cal and Dale:

Art

Cal

Bill

Dale

Conversely, suppose that Cal was at least as tall as Art:

Art Cal

Bill

Dale

The difference between Art and Bill would now be smaller than the difference between Cal and Dale, contrary to the first premise. Even if you start by building a model of the first premise, the inference still depends on the relation between the relations. Hence, we refer to this sort of problem as *fully-relational*. The model theory predicts that these problems should be more difficult than single-clause problems, because their solution depends

on all the information in the models of both premises. This difference should arise whether or not participants rely on *external* or diagrammatic models of the premises, or, alternatively, *internal* mental models.

All problems with two premises that refer to four individuals and that yield a single consistent ordering are of the two sorts that we have illustrated. Given a series of inferences that are similar to each other, reasoners tend to develop a strategy to cope with them (see, e.g., Van der Henst, Yang, & Johnson-Laird, 2002). One strategy for the present problems, as we previously outlined, is to order the three individuals referred to in the second premise and then to integrate the remaining information. Another strategy would be to infer a transitive order from separate premises. For example, in the preceding problem the first clause of the first premise, Art is taller than Bill, could be combined with the second clause of the second premise, Bill is taller than Dale, to infer the order:

Art

Bill

Dale

This model could then be integrated with the remaining information to produce the final rank order: The first premise establishes that Art is taller than Cal, and the second premise establishes that Cal is taller than Bill. Hence, Cal can be interpolated between Art and Bill.

Yet another strategy would be to represent the premises strictly in their sequential order. That is, reasoners would build a model of the first premise and then add to it any information from the second premise that was required to solve the problem.

Whatever strategies reasoners use, they can solve these fully-relational problems only by taking into account all the information in the first premise—both relations and the relation between these relations. In contrast, they can solve single-clause problems by taking into account only a single relation described in one clause of the first premise. It follows that

single-clause problems should be easier than fully-relational problems. This account locates the difficulty of these sorts of problem in the process of integrating the information in the two premises (henceforth, the *integration* process). To develop a better understanding of how individuals might reason, we wrote a computer program (in LISP) that could solve all 600 possible two-premise problems based on relations between binary relations. The program itself was primarily an exercise in artificial intelligence. It constructed a list of all possible orders of the individuals and used the information in the premises to eliminate members from the list. This overall strategy is most implausible psychologically, and we will not describe the details of the program. However, component parts of the program were psychologically informative. The program showed that the crucial step in relational reasoning is likely to be the integration of models of the two premises rather than the initial complexity of building models of the individual premises (because integration required a far more sophisticated code than did model building).

EXPERIMENT 1A

All problems with a single consistent ordering are either single-clause or fully-relational problems. The model theory predicts that the single-clause problems should be easier than the fully-relational problems. This difference should occur in the accuracy of responses, or, failing that, in their latencies. The aim of our first experiment was to test these two predictions.

Method

Participants

A total of 16 students (9 male, 7 female) from Princeton University received course credit for their participation in the experiment.

Design

The participants acted as their own controls and carried out four single-clause and four

fully-relational problems. The first premise was of the same form throughout, but the second premise varied depending on the sort of problem. The forms of the eight problems are shown in Appendix A. We also used four problems that had no consistent rank-order solution as filler items. The problems were presented in a different random order to each participant.

Materials

The problems were based on three different transitive and asymmetrical relations: *taller than*, *heavier than*, and *bigger than*. We made three different assignments of these relations to the eight forms of problem. The first assignment was at random, and the remaining assignments were made by shifting each relation to the next problem in the set. Each participant was tested with one of the resulting sets selected at random. We constructed eight different sets of four 1-syllable names, four sets of female names, and four sets of male names. The names within a set had different initial letters, and none of them rhymed. Each participant received a new random assignment of the set of names to the eight problems.

Procedure

The participants were tested individually and carried out the experiment using paper and pencil. The instructions were presented on an initial sheet of paper, and the key instructions were as follows:

Given that both premises are true, your task is to decide on a possible ordering of the four people in terms of the property in question. Please always write the tallest, or biggest, or heaviest, person at the top of your list. If, however, you think that there is no possible ordering consistent with the premises, then write down: "none". Please take as much time as you need to complete each problem, keeping in mind that we will be timing how long it takes you to do each problem.

Each problem was presented on a separate sheet of paper. The two premises were at the top of the

page, under which there was a large area for written working. At the bottom of each page there was a box in which the participants wrote down their final answer. They were told:

... You may make workings on the paper above this box, and then when you have made a final decision on your answer, enter it into the box. Once you have made an entry in the box, you cannot go back, and must move onto the next problem.

The participants' responses were timed from the moment at which they received the problem until they started to write their response in the box.

Results and discussion

The percentages of correct conclusions for the two sorts of problem and the mean latencies of all the responses, correct and incorrect, were as follows:

| | |
|----------------------------|-----------|
| Single-clause problems: | 83%, 47 s |
| Fully-relational problems: | 88%, 56 s |

There was no reliable difference in the accuracy of performance (Wilcoxon test, $z = 0.58$, $p > .5$, two-tailed). However, single-clause problems were solved more rapidly than fully-relational problems, and the result was reliable both for all responses, correct and incorrect (Wilcoxon test, $z = 1.81$, $p < .05$, one-tailed), and for the correct responses alone (Wilcoxon test, $z = 2.53$, $p < .01$, one-tailed). Hence, the amount of information to be integrated to form a unitary model affects performance. The process is faster for a single binary relation than for a relation between two binary relations. These results are compatible with the notion that reasoners rely on mental models to solve such problems—that is, that the difficulty of integrating the two premises arises from difficulties in combining mental models of the two premises. However, they do not decisively support this account. Our participants used paper and pencil, but the observed difference should also occur if individuals do not rely on external diagrams.

EXPERIMENT 1B

We carried out a second experiment in which participants were not permitted to make any written workings before writing their final response. If participants rely on mental models when they cannot draw them as diagrams, as in Experiment 1a, then the experiment should corroborate the prediction that single-clause problems are easier than fully-relational problems.

Method

Participants

A total of 18 participants (7 male, 11 female) from Princeton University participated for course credit.

Design

The participants acted as their own controls and carried out the same four single-clause and four fully-relational problems as those used in Experiment 1A. The first premise was of the same form throughout, but the second premise varied depending on the sort of problem. All eight problems, which are shown in Appendix A, yielded a consistent ordering. Each participant received the problems in a different random order.

Materials

As in Experiment 1A, the problems were based on three different transitive relations: *taller than*, *heavier than*, and *bigger than*. We used the same procedure as Experiment 1A to assign these relations to the eight forms of problem and to assign the eight different name sets. Each participant was tested with one of the resulting three sets selected at random.

Procedure

The participants were tested individually. The general procedure and the key instructions were the same as those in the first experiment, except that the participants were not allowed to make any written workings. They had to solve each problem mentally. The participants' responses were timed from the moment at which they

received the problem until they started to write their response in the box.

Results

The percentages of correct conclusions for the two sorts of problem and the mean latencies of all the responses, correct and incorrect, were as follows:

Single-clause problems: 82%, 58 s

Fully-relational problems: 76%, 76 s

The single-clause problems did not yield a reliably greater proportion of correct responses than the fully-relational problems, although the trend was in that direction (Wilcoxon test, $z = 0.81$, $p > .2$, one-tailed). However, the overall latencies for correct and incorrect responses were reliably faster for the single-clause problems than for the fully-relational problems (Wilcoxon test, $z = 3.07$, $p < .01$, one-tailed). This difference was also reliable in a separate analysis of the correct responses alone (Wilcoxon test, $z = 2.87$, $p < .01$, one-tailed). Hence, as the theory predicts, the participants dealt with the single-clause problems faster than the fully-relational problems.

This replication corroborates the account of the difficulty of the problems based on the amount of information to be integrated to form a unitary model. The greater difficulty of fully-relational problems is not based on their need for more complicated external diagrams. The results therefore go further than Experiment 1A in supporting the idea that reasoners rely on mental models to solve these relational problems. However, the support is not decisive, because other theories may also be able to account for this difference (see General Discussion).

The results of both Experiments 1A and 1B suggest that reasoners do not tackle these problems by representing the premises strictly in their sequential order. The fully-relational problems had two sorts of second premise yielding an ordering of three individuals. One sort was of the form $(C > B) > (B > D)$, which yields a transitive ordering. Another sort was of the form $(C > D) > (B > D)$, which calls for the relation between the

two clauses to be processed in order to yield the ordering. We have independently demonstrated that when considered by themselves, the first sort of premise yields faster inferences than does the second sort of premise (Goodwin & Johnson-Laird, 2005). Hence, if reasoners were processing the premises sequentially, then they should solve problems with the first sort of second premise, $(C > B) > (B > D)$, faster than they solve those with the second sort of second premise, $(C > D) > (B > D)$. Yet, this difference did not occur. In fact, the trend in both experiments was in the opposite direction (Experiment 1A, Wilcoxon test, $z = 1.24$, $p > .2$, two-tailed; Experiment 1B, Wilcoxon test, $z = 1.85$, $p > .06$, two-tailed). Hence, reasoners seemed not to be building a model of the first premise and then adding information from the second premise, but to be relying instead on a different strategy. To examine this possibility and to elucidate the mental processes for reasoning about relations between relations, we carried out the next experiment.

EXPERIMENT 2

To discover the strategies that individuals develop, the participants had to think aloud as they tackled single-clause and fully-relational problems (see, e.g., Ford, 1995; Van der Henst et al., 2002). The analysis of such protocols is labour intensive, and so we tested a smaller number of individuals than usual. Our aim was to characterize the strategies that individuals spontaneously developed and to check that they were genuinely reasoning as opposed to guessing in a constrained way. An analysis of the participants' strategies allows a more thorough investigation of individual differences and of the fundamental processes underlying reasoning (Roberts, 2000).

Method

Participants

A total of 11 participants (5 male, 6 female) from Princeton University were paid \$8 for their participation in the experiment.

Design

The participants acted as their own controls and carried out the same four single-clause and four fully-relational problems as those used in Experiment 1. The first premise was of the same form throughout, but the second premise varied depending on the sort of problem. All eight problems, which are shown in Appendix A, yield a consistent ordering. Each participant received the problems in a different random order.

Materials

As in Experiment 1, the problems were based on three different transitive relations: *taller than*, *heavier than*, and *bigger than*. We used the same procedure as that in Experiment 1 to assign these relations to the eight forms of problem and to assign the eight different name sets. Each participant was tested with one of the resulting three sets selected at random.

Procedure

The participants were tested individually and carried out the experiment using paper and pencil. The general procedure and the key instructions were the same as those in the first experiment, but the participants were also instructed to think aloud. They had to say all the thoughts that occurred to them while they worked through each problem, and we video-recorded what they wrote, drew, and said. The video camera was on a tripod standing on the desk in front of them. It was angled down towards the page on which they were working, so that the whole response sheet was recorded. The experimenter sat behind the participants and watched their responses on a colour monitor that was not visible to them. Whenever anyone fell silent for more than three seconds, the experimenter prompted them to think aloud. The participants' responses were timed from the moment at which the problem sheet was put in front of them until they started to write their response in the answer box.

Results

The percentages of correct conclusions for the two sorts of problem and the mean latencies of all the responses, correct and incorrect, were as follows:

Single-clause problems: 93%, 59 s

Fully-relational problems: 86%, 88 s

The single-clause problems did not yield a reliably greater proportion of correct responses than the fully-relational problems, although the trend was in that direction (Wilcoxon test, $z = 1.13$, $p > .1$, one-tailed). However, the overall latencies for correct and incorrect responses were reliably faster for the single-clause problems than for the fully-relational problems (Wilcoxon test, $z = 2.31$, $p < .02$, one-tailed). This difference was also reliable in a separate analysis of the correct responses alone (Wilcoxon test, $z = 2.05$, $p < .03$, one-tailed). Hence, as the theory predicts, the participants dealt with the single-clause problems faster than the fully-relational problems.

Strategies and tactics

We transcribed the protocols and were able to make sense of the strategies that the participants used on 92% of the problems.¹ At first, they tended to flail around and were uncertain how to proceed, but even on the first problem they were almost always able to arrive at a correct solution. In some cases, they abandoned one strategy and took up another in order to solve a problem—in such cases, we considered only the second successful strategy in the following analysis. In general, the participants worked through the problems in a sequence of steps, but sometimes they would cycle through the same step more than once (see also Van der Henst et al., 2002). On occasion, they omitted a step altogether. As in the previous experiment, reasoners almost always read the first premise first, but the strategy that they used to solve the problem more often started with a full representation of the second premise.

¹A complete record of each of the protocols in Experiment 2 can be found on the second author's webpage: <http://webscript.princeton.edu/~psych/PsychSite/suppstrats.doc>

All the participants began by drawing a diagram, by writing an expression with “>”, or by writing a numerical expression to represent the first premise. Their next step was to read the second premise. They constructed an initial ordering of three individuals (on 75% of trials) or two individuals (on 13% of trials), but sometimes they constructed a final ordering of all four individuals (on 8% of trials). Most of the orderings of three individuals were derived from the second premise (65%), but some were derived from transitive inferences using clauses from separate premises (32%). Most of the orderings of two individuals were based on a single clause (67%), but the remaining orderings began with a “frame” based on the hypothesized tallest and shortest individuals. After the participants had constructed an initial ordering, their next step, where necessary, was to produce an integrated ordering of all four individuals. Typically, they made a final check of the ordering against the two premises. All of these facts about the protocols are obvious in them.

Our evidence is consistent with the idea that the participants relied on models, although there are certain caveats with this interpretation (see General Discussion). One participant made an initial use of algebraic expressions, but he subsequently abandoned this procedure in favour of using models. Within the possible strategies based on models, there are three a priori possibilities, and, in fact, we observed each of them:

1. The *ternary* strategy (49% of observed strategies): Reasoners used one premise to construct an ordering of three individuals (only the second premise in our problems afforded this information). They then found the relevant information from the other premise to place the fourth individual in the ordering. Table 1 shows a verbatim protocol of a participant using this strategy. As readers will see, the protocol shows the sequence of steps with little or no ambiguity.
2. The *transitive* strategy (24% of observed strategies): Reasoners mentally conjoined a diagram (or model) of a binary relation with a diagram of another binary relation

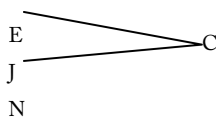
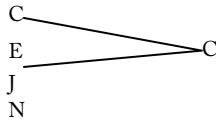
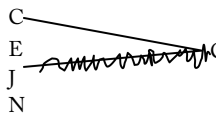
in order to yield an integrated transitive order of three individuals. Each binary relation came from a separate premise. They then imported the fourth individual into the ordering from the second premise. Table 2 shows a verbatim protocol of a participant using this strategy.

3. The *frame* strategy (17% of observed strategies): Reasoners constructed an initial frame of one or two individuals, by typically inferring, say, the tallest individual and using the premises to make a plausible inference about who was then the shortest. The tallest individual was often identified by observing that at least within the four clauses of the problem, no-one was described as taller than that particular individual. On occasion, reasoners would instead use a model of one of the clauses from the premises as the frame. They would then insert the remaining individuals within the frame using an educated guess. Table 3 shows a verbatim protocol of a participant using this strategy.

In general, the protocols of the three strategies were fairly unambiguous about the sequences of steps that the participants took. The remaining protocols were either misinterpretations of the premises (2%) or not sufficiently clear for us to discern a definite strategy (9%).

Most of the responses were correct, and the strategies did not differ significantly in the accuracies or latencies of the responses that they yielded. However, there was strategic variation both within and between participants. The participants used a mean of 2.27 out of the three main strategies; one participant, however, used the same strategy throughout. If a participant used, say, two strategies, then he or she was bound to shift strategy at least once. The participants, however, made a mean of 2.9 strategy shifts (compared with a mean of 1.27 necessary strategy shifts given the number of observed strategies). They shifted more often than necessary (Wilcoxon test, $z = 2.42$, $p < .02$, two-tailed). The ternary strategy was the modal strategy for 7 participants,

Table 1. *A verbatim protocol of a participant using the ternary strategy Problem. Chad is heavier than John to a greater extent than Ed is heavier than Nick. Ed is heavier than Nick to a greater extent than John is heavier than Nick.*

| <i>Verbal protocol</i> | <i>Diagrams drawn by the participant</i> | |
|---|--|--|
| C is heavier than J to a greater extent than E is heavier than N. | C J | E N |
| E is heavier than N to a greater extent than J is heavier than N. | | |
| So E, N, | | E (new diagram) N |
| J's going to be somewhere in there. | | E J N |
| C is heavier than J to a greater extent than E is heavier than N. | | E J N C |
| E is heavier than N to a greater extent than J is heavier than N. | | |
| C is going to go here or here. | |  |
| C is heavier than J to a greater extent than E is heavier than N. | | |
| E is heavier than N to a greater extent than J is heavier than N. | | |
| C is heavier than J. | | |
| C is heavier than J to a greater extent than E is heavier than N. | | |
| So we can only assume that C is going to be here. | |  |
| C is heavier than J to a greater extent than E is heavier than N. | | |
| So I'd say that's where it's going to be. | |  |
| So it's going to be C, E, J, N. | | |

Note: The convention for these tables is that consecutive diagrams are part of the same overall diagram, except where a new diagram is indicated in parentheses. Participants spoke the full names where we have represented them with single letters.

Conclusion: Chad, Ed, John, Nick (correct).

General comment: The participant represents the clauses of the first premise in a diagram. She then constructs an ordering of three individuals using the information in the second premise. She then uses the fully-relational information in the first premise to place the remaining fourth individual at the top of this ordering. She accepts this ordering as correct without further checking.

Table 2. *A verbatim protocol of a participant using the transitive strategy Problem. Steve is bigger than Doug to a greater extent than Bob is bigger than Ray. Bob is bigger than Ray to a greater extent than Bob is bigger than Steve.*

| Verbal protocol | Diagrams drawn by the participant | | | |
|---|-----------------------------------|---|---|--|
| S is bigger than D. | S | | | |
| | D | | | |
| And B is bigger than R. | S | B | | |
| | D | R | | |
| Now B is bigger than S. | S | B | B | |
| | D | R | S | |
| So this goes there. | B | | | |
| | S | B | B | |
| | D | R | S | |
| Now, B, S, D. | | | | |
| And between B and R and B and S.ok. | | | | |
| B is bigger than R to a greater extent than B is bigger than S. | B | | | |
| | S | B | B | |
| | — R | | | |
| | D | R | S | |
| B, S, R, D | | | | |

Note: The convention for these tables is that consecutive diagrams are part of the same overall diagram, except where a new diagram is indicated in parentheses. Participants spoke the full names where we have represented them with single letters.

Conclusion: Bob, Steve, Ray, Doug (correct).

General comment: The participant starts by representing both clauses of the first premise. He then conjoins a model of the second clause from the second premise (Bob is bigger than Steve) with a model of the first clause of the first premise (Steve is bigger than Doug) to construct an ordering of three individuals: Bob, Steve, Doug. He uses the second premise to infer that Ray is smaller than Steve because Bob is bigger than Ray to a greater extent than Bob is bigger than Steve, and he then imports Ray into the previous ordering of three individuals. Here, he overlooks that he has recovered no definite information about the relative heights of Ray and Doug. In general terms, he is using the second premise to infer an ordering of three individuals (Bob, Steve, and Ray), and imports one of them (Ray) into an existing ordering of three individuals (Bob, Steve, and Doug). He finishes without checking his final ordering against the premises.

Table 3. *A verbatim protocol of a participant using the frame strategy Problem. Flo is heavier than Mel to a greater extent than Jan is heavier than Sky. Jan is heavier than Flo to a greater extent than Jan is heavier than Sky.*

| Verbal protocol | Diagrams drawn by the participant | |
|---|--------------------------------------|--|
| F is heavier than M to a greater extent than J is heavier than S. | $F \rightarrow M > J \rightarrow S.$ | |
| J is heavier than F to a greater extent than J is heavier than S. | | |
| So J must be the heaviest, followed by.let's see, M would probably be the smallest. | J (new diagram) | |
| | M | |
| So then, J is heavier than F to a greater extent than J is heavier than S. | | |
| So that would put S here. . . | J S | |
| | M | |
| and F here. | J S F M | |
| F is heavier than M to a greater extent than J is heavier than S. | | |
| J is heavier than F to a greater extent than J is heavier than S. | | |

Note: The convention for these tables is that consecutive diagrams are part of the same overall diagram, except where a new diagram is indicated in parentheses. Participants spoke the full names where we have represented them with single letters.

Conclusion: Jan, Sky, Flo, Mel (correct).

General comment: The participant starts by representing the first premise using arrows to represent "heavier than". She then constructs a frame of two individuals, who she supposes are the heaviest and the lightest (Jan and Mel). Within this frame she imports an ordering of three individuals derived from the second premise. This step creates the final ordering, which she checks against both premises.

the frame strategy was the modal one for 3 participants, and the transitive strategy was the modal one for 1 participant.

A strategy, in effect, is based on a set of tactical steps, and it constrains the order in which these tactics can occur. So, it is sensible to ask what the possible tactical steps are. In fact, there are eight a priori possible tactics out of which a strategy can be constructed. Six of these tactics exploit binary relations, and two of them exploit relations between relations. In what follows, we describe each tactic. The six tactics using binary relations are:

1. Add a binary relation. Take a binary relation and use it to add an item to a model of an ordering:

| | | | | |
|---|---|---|---|---|
| A | + | B | ⇒ | A |
| B | | C | | B |
| | | D | | C |
| | | | | D |

The tactic is valid. It was used in the transitive strategy to conjoin clauses from separate premises. It was also used to place the fourth individual in the single-clause problems.

2. Combine *n*-ary relations. Take models of two *n*-ary orders and integrate them:

| | | |
|---|---|---|
| A | B | A |
| B | + | D |
| C | | C |
| | | D |
| | | C |

Although this tactic is similar to the previous one, there is a difference. It is not always valid, as in the following example:

| | | |
|---|---|---|
| A | A | A |
| B | + | B |
| C | | D |
| | | D |
| | | C |

It was employed most frequently in the transitive strategy and occasionally in the frame strategy.

3. Build a frame (i.e., a model of the relation between at least two individuals), by identifying one end item, such as the tallest or shortest individual, and then the other.

For instance, given the following sort of problem:

(A > B) > (C > D)
 (C > D) > (C > A)

one participant set out the four clauses vertically:

| | | | |
|---|---|---|---|
| A | C | C | C |
| B | D | D | A |

and reasoned: "C is the tallest individual in three out of four clauses, and no-one is taller than C in any other clauses, therefore C is the tallest." Another participant reasoned: "D is shortest in two out of four clauses (i.e., more than anyone else), and no-one is shorter than D in any other clauses, therefore D is the shortest." These two sorts of tactic are not valid, but for this problem the inference of the tallest individual is correct, whereas the inference of the shortest is not. These two tactics were used in the frame strategy.

4. Build a frame by selecting a single clause at random, say, C > D, and use it as an initial ordering:

C
D

This tactic was also used in the frame strategy.

5. Constrained relational guessing:

| | | | |
|---|---|---|---|
| A | C | ⇒ | A |
| B | B | | C |
| | | | B |

This tactic was used in the frame strategy to place individuals within, or occasionally outside, the initial frame. When reasoners made guesses, they would typically either corroborate or revise their guesses by checking them against the premises. Although it was difficult to determine what reasoning was occurring during checking (participants would typically just read aloud the premises), our

analysis suggests that they must have been employing one or other of the tactics we set out here. Corroboration did not rely on a new sort of tactic.

6. Unconstrained guessing:

$$\begin{array}{ccc} A & B & \Rightarrow & A \\ & & & B \end{array}$$

Although this tactic was an a priori possibility, it was rarely observed, because guessing mainly occurred using an initial ordering as a constraint.

There were two tactics based on relations between relations:

7. Construct a ternary order. Construct a model of a ternary order from a single premise, for example $(B > D) > (C > D)$:

$$\begin{array}{ccccc} B & | & & & B \\ & & | & C & \Rightarrow & C \\ D & | & D & & & D \end{array}$$

This tactic was the main step in the ternary strategy.

8. Combine fully-relational information with a model of an n -ary order:

$$\begin{array}{ccccccc} A & | & & & & & A \\ & & | & & & & \\ & & | & C & + & C & \Rightarrow & C \\ & & | & D & & B & & B \\ B & | & & & & D & & D \end{array}$$

This tactic was needed only for fully-relational and not for single-clause problems.

We observed all of these tactics in the protocols. Their frequency of occurrence depended on the strategy in use and on the complexity of the problem. For instance, the construction of a model of a ternary order (Tactic 7) occurred only when individuals used the ternary strategy, and the combination of fully-relational information with a ternary order (Tactic 8) occurred only for fully-relational problems. Each of the three main

strategies depended on distinct combinations of these tactics. However, there was also some variation within each strategy in the selection and sequence of tactics—not every instance of a particular strategy was identical.

EXPERIMENT 3

In the previous experiment, the large variation within participants in their use of strategies suggested that they were experimenting with various tactical steps and assembling strategies “bottom up” from sequences of these steps. The model theory postulates that various non-logical characteristics of problems given to participants should influence which particular strategies they develop (Van der Henst et al., 2002). More specifically, reasoners ought to approach the same problem differently if, in two different presentations, problem information enters working memory in a different order, or different aspects of the problem are made more salient. Experiment 3 tested this prediction.

One way in which to elicit a particular reasoning strategy should be to make salient the information required in the initial steps of the strategy. The ternary strategy takes information from a single premise and builds an ordering of three individuals. In the first two experiments, this information was always provided in the second premise. Hence, one way in which to control the use of the ternary strategy should be to manipulate premise order—reasoners should be more likely to develop the ternary strategy when the first premise yields an ordering of three individuals. The transitive strategy conjoins binary relations from separate premises. If the clauses yielding these relations are made more salient (by colour coding, say), reasoners should be more likely to develop the transitive strategy. The ternary strategy ought theoretically to be more efficient than the transitive and frame strategies. Experiment 2 revealed a tendency in that direction, but the experiment may not have been powerful enough to yield a reliable result. Nevertheless, once individuals have developed

the ternary strategy, they should be less likely to abandon it than to abandon the transitive or frame strategies. Experiment 3 tested these three predictions.

Method

Participants

A total of 18 participants (7 male, 11 female) from Princeton University were paid \$8 for their participation in the experiment.

Design

The participants acted as their own controls and carried out eight problems: four single-clause and four fully-relational problems. Each participant received two blocks of four problems. Each block consisted of two single-clause and two fully-relational problems, and these same four problems were used in both blocks (with different names and relations in the second block). Hence, there were only four unique sorts of problem. The blocks were designed to inculcate either the ternary strategy or else the transitive strategy. Half the participants received the ternary block first and then the transitive block, and half the participants received the two blocks in the opposite order. The first premise varied depending on the block. In the ternary block, the first premise yielded an ordering of three individuals, whereas in the transitive block the premise yielding this information was second. In the transitive block, the two premises that yielded the transitive inference were printed in red (the remainder of the text was printed in black). The full set of problems, all of which yielded a single consistent ordering, is shown in Appendix B. Within each block, each participant received the four problems in a different random order.

Materials

As in the first two experiments, the problems were based on three different transitive relations: *taller than*, *heavier than*, and *bigger than*. We used the same procedure as that in the previous experiments to assign these relations to the eight forms of problem and to assign the eight different name

sets. Each participant was tested with one of the resulting three sets selected at random.

Procedure

The participants were tested individually and carried out the experiment using paper and pencil. Participants were instructed to think aloud. They had to say all the thoughts that occurred to them while they worked through each problem, and what they said, wrote, and drew was video-recorded. The general procedure and instructions were the same as those in Experiment 2.

Results and discussion

The percentages of correct conclusions for the two sorts of problem and the mean latencies of all the responses, correct and incorrect, were as follows:

| | |
|----------------------------|-----------|
| Single-clause problems: | 86%, 72 s |
| Fully-relational problems: | 69%, 94 s |

The single-clause problems yielded a greater proportion of correct responses than did the fully-relational problems (Wilcoxon test, $z = 2.04$, $p < .03$, one-tailed). Participants responded faster to the single-clause problems than to the fully-relational problems, and the difference was reliable both overall (Wilcoxon test, $z = 2.85$, $p < .01$, one-tailed) and for the correct responses alone (Wilcoxon test, $z = 2.5$, $p < .01$, one-tailed). These latency results replicated the findings from the first two experiments and corroborated the present theory. Individuals are faster in integrating the premises for single-clause problems than for fully-relational problems. They also appear to be slightly more accurate in reasoning with single-clause problems. We emphasize the difference observed in accuracy in Experiment 3 (which was not reliable in any of the previous three experiments) for two main reasons. First, the same trend was observed in Experiment 2, although that experiment may not have had sufficient power to detect a real difference. Second, the analysis of strategies shows that it is possible to arrive at the correct conclusion

for the wrong reasons. In both Experiments 2 and 3, participants sometimes used invalid tactics without paying the price of an incorrect conclusion on the fully-relational problems (particularly when they used the transitive or frame strategies). The requirement to think aloud may have diminished this tendency somewhat by increasing vigilance. Hence, the lack of an accuracy difference in Experiments 1a and 1b may have resulted from less careful reasoning on the part of our participants. It both decreased performance on the single-clause problems (82% averaged across Experiments 1a and 1b vs. 89% averaged across Experiments 2 and 3) and artificially increased performance on the fully-relational problems (82% in Experiments 1a and 1b vs. 76% averaged across Experiments 2 and 3). This difference was marginally reliable (Wilcoxon test, $z = 1.83$, $p < .07$), indicating that thinking aloud did affect performance. An analysis of the experiments in which written working was permitted (i.e., dropping Experiment 1b) showed that the difference attributable to thinking aloud is reliable (Wilcoxon test, $z = 2.13$, $p < .04$). We regard the accuracy data in Experiments 2 and 3 as providing a better account of the true difficulty of the problems.

We transcribed the protocols, and we were able to make sense of the strategies that the participants used on 92% of the problems. As in Experiment 2, reasoners typically exhibited some initial uncertainty and would cycle through the same step more than once, but even on the first problem they were often able to arrive at a correct solution. In some cases, participants switched from one strategy to another. The participants used the same three strategies that we observed in Experiment 2. The overall percentages were as follows:

Ternary strategy: 40
 Frame strategy: 25
 Transitive strategy: 17

As in Experiment 2, the protocols of these strategies are fairly unambiguous, and their percentages are roughly in line with that experiment, though the frame strategy was more prevalent in the present experiment. The remaining protocols

were mixtures of two strategies (4%), partial use of one of the main strategies (5%), or not sufficiently clear for us to discern a strategy (8%).

Table 4 presents the percentages of use of each of the three main strategies in the two blocks. Overall, the participants' strategies were biased by the condition. The condition inculcating the ternary strategy yielded this strategy on 54% of trials, whereas the condition inculcating the transitive strategy yielded the ternary strategy on only 26% of trials (Wilcoxon test, $z = 2.57$, $p < .01$, one-tailed). The transitive condition, in fact, tended to produce either the transitive or the frame strategy. Evidently, the emphasis on the two transitive clauses encouraged the participants to identify end items in the ordering. Some participants in this condition, without explicitly drawing the relevant transitive inference, may have been aware of the possibility of this inference. They could then select one of the end items in the transitive ordering as an end element for a frame.

The clearest indication of the inculcation of strategies occurred in the first block of trials—the participants in the ternary condition used the ternary strategy more often than did participants in the transitive condition (Mann–Whitney

Table 4. *The percentages of strategies in the two blocks of Experiment 3 depending on which strategy was inculcated*

| | Strategy used | Strategy inculcated | |
|--------------|---------------|---------------------|------------|
| | | Ternary | Transitive |
| First block | Ternary | 53 | 6 |
| | Transitive | 14 | 17 |
| | Frame | 8 | 53 |
| Second block | Ternary | 56 | 47 |
| | Transitive | 14 | 25 |
| | Frame | 22 | 17 |
| Overall | Ternary | 54 | 26 |
| | Transitive | 14 | 21 |
| | Frame | 15 | 35 |

Note: The participants whose first block inculcated the ternary strategy received a second block inculcating the transitive strategy, and vice versa. The percentages do not sum to 100: The remainder of the trials were combinations of strategies, partial instances of a strategy, or unclear.

$U = 17.5$, $z = 2.22$, $p < .02$, one-tailed). There was evidently transfer from the first block of trials to the second block. Once the participants had developed the ternary strategy in the first block, they tended to continue to use it in the second block. As Table 4 shows, participants who received the ternary condition in the first block showed a relatively small decline in the use of the strategy in the second block, whereas participants who received the transitive condition in the first block showed a large decline in the use of the other two strategies (transitive and frame) in the second block. Hence, the ternary strategy transferred to a greater degree than the other strategies (Mann-Whitney $U = 17.0$, $z = 2.22$, $p < .02$, one-tailed).

Although the participants did not use any new strategies, they did sometimes use a more advanced version of the frame strategy than in Experiment 2. When they had identified, say, the tallest individual, they would lay out the relative differences between this individual and the others by drawing lines as a visual aid. They then used these lines to arrange the remaining individuals in order. Table 5 shows a verbatim protocol of a participant using this strategy. The strategy was accurate provided that the end individual was identified correctly, but this identification relied on constrained guesswork.

As in the previous experiment, there was both between- and within-subject variation in the use of strategies. The ternary strategy was modal for 9 participants, the frame strategy was modal for 5 participants, and the transitive strategy was modal for 2 participants. One participant used the transitive and frame strategies equally often, and another participant used the ternary and frame strategies equally often. The participants used a mean of 2.06 out of the three main strategies; three participants, however, used the same strategy throughout. As we described earlier, if participants use, say, two strategies, then they are bound to shift strategies at least once. Our participants made a mean of 1.56 strategy shifts (compared with a mean of 1.06 necessary strategy shifts given the number of observed strategies). Unlike in Experiment 2, they did not shift reliably more often than necessary (Wilcoxon test,

$z = 0.34$, $p > .3$, one-tailed). This diminished variation in the use of strategies provides indirect evidence that the biasing manipulations did constrain the participants' strategies.

There were no robust signs of differences in accuracy between the three strategies, but a trend towards less accurate performance with the frame strategy (83% correct with ternary strategy; 88% correct with the transitive strategy; and 69% correct with the frame strategy). However, the strategies did appear to differ in the times that they took participants. The mean latencies (seconds) for all responses (left), and for correct responses only (right) were as follows:

| | | |
|----------------------|------|-----|
| Ternary strategy: | 66; | 62 |
| Transitive strategy: | 84; | 88 |
| Frame strategy: | 108; | 110 |

We examined the differences treating each trial as independent of the others, and they were marginally reliable both for responses overall, Friedman test, $\chi^2(2) = 6$, $p < .06$, and for the correct responses only, Friedman test, $\chi^2(2) = 4.67$, $p < .1$. There was a trend for the ternary strategy to be used more frequently on the fully-relational than on the single-clause problems. A total of 7 out of 18 participants demonstrated this pattern, and only one showed the reverse pattern (Wilcoxon test, $z = 2.13$, $p < .02$, two-tailed).

We considered the possibility that certain fully-relational problems may be more difficult than others. For all fully-relational problems, one premise yields an ordering of three individuals, to which a fourth individual needs to be added. We tested whether problems in which the fourth individual needed to be added to the bottom of the ordering differed from those in which the fourth individual needed to be added to the top (see Huttenlocher, 1968; Huttenlocher & Strauss, 1968). However, even a combination of the data from all four studies failed to yield a reliable difference between the two sorts of problem in terms of accuracy, overall latencies, or latencies for correct responses only (Wilcoxon tests, $z = 1.3$, $z = 0.89$, $z = 0.58$, respectively, p s $> .15$, in all cases, two-tailed).

Table 5. *A verbatim protocol of a participant using a more advanced frame strategy in Experiment 3. Problem: Val is taller than Jill to a greater extent than May is taller than Brooke. May is taller than Brooke to a greater extent than May is taller than Val.*

| <i>Verbal protocol</i> | <i>Diagrams</i> |
|---|---|
| V is taller than J to a greater extent | V J |
| V is taller than J to a greater extent | V ↑ (new diagram) ↓ J |
| V is taller than J to a greater extent than M is taller than B. | |
| So, M is taller than B to a lesser extent. . . . | V ↑ ↓ J M ↔ B ↔ |
| M is taller than B to a greater extent. . . . | |
| M is taller than B to a greater extent than M is taller than V, so M is taller than V. | M ↑ (new diagram) ↓ B M ↔ V ↔ |
| So I guess V is taller than J. | |
| V is shorter than M, so I have | |
| M is taller than V, so M is the biggest one I suppose. | M (new diagram) |
| And I have M is taller than B, M is taller than B, yes | |
| M is taller than V . . .to a lesser extent . . . what about J? . . .doesn't say much about J so I don't . . . | |
| Alright M is the tallest. | |
| M is taller than B to a greater extent. . .so | |
| Let's see . . . V is shorter than B? . . let's see | |
| All I know is that M is taller than Bok. | |
| And M is taller than V. | |
| But M being taller than B is greater than M being taller than V. . .so M must be the second biggest (this step was accompanied by the participant pointing to the arrows that were drawn in the model of the second premise). | M V |

(Continued overleaf)

Table 5. (Continued)

| <i>Verbal protocol</i> | <i>Diagrams</i> |
|---|------------------|
| And I know that V is bigger than J . . . does it compare J to B? . . .let's see. . . M is taller than B . . . yeah so B is shorter than V, and J, so B must be the third (this step was accompanied by the participant pointing to the arrows drawn in the model of the first premise). | M V B |
| And J must be the last one. | M V B J |

Note: The convention for these tables is that consecutive diagrams are part of the same overall diagram, except where a new diagram is indicated in parentheses. Participants spoke the full names where we have represented them with single letters.

Conclusion: May, Val, Brooke, Jill (correct).

General comment: The participant starts by representing the first premise using a diagram, with arrows to represent the relative differences in height between the individuals in each clause. He represents the second premise in the same way. He then makes a constrained guess that May is the tallest individual, and May serves as the frame. By comparing the relative length of the lines in the model of the second premise, the participant concludes that Val is the second tallest individual and accordingly places Val beneath May. This partial ordering of two individuals is then considered together with the length of the lines in the model of the first premise. This results in the remaining two individuals being placed at the bottom of the ordering.

GENERAL DISCUSSION

Our aim was to understand the mental processes underlying reasoning about relations between relations. Such reasoning occurs in many domains, from daily life to science (see the Introduction). To ensure that individuals reasoned about such higher order relations, we gave them two main sorts of problem in which they had to put four individuals into their appropriate rank order. With *single-clause* problems, such as:

Rich is taller than Neil to a greater extent than Joe is taller than Dale.

Neil is taller than Dale to a greater extent than Joe is taller than Dale.

they could use the second premise to envisage the following ternary relation in height (shown vertically):

Neil

Joe

Dale

In this case a single clause from the first premise, Rich is taller than Neil, yields a binary relation that establishes the overall rank order:

Rich

Neil

Joe

Dale

In *fully-relational* problems, however, such as:

Rich is taller than Joe to a greater extent than Neil is taller than Dale.

Neil is taller than Dale to a greater extent than Joe is taller than Dale.

reasoners need to use the relation between the two relations in the first premise in order to solve the problem. Because the integration of information from both premises into a single model is the most complex step in the task, as shown by a computer program that solved the problems, it should be the prime contributor to the task's *effective relational complexity* (see Halford et al., 1998). Hence, the single-clause problems should be easier than the fully-relational problems.

Experiments 1a and 1b supported this prediction for the latencies of response, but the difference in accuracy between the two sorts of problem was not reliable. A further analysis of the latencies for the fully-relational problems

suggested that participants did not necessarily represent the premises in the order in which they were presented. Experiment 2 examined the same problems, and the response latencies again corroborated the predictions of the theory. Experiment 3 corroborated the prediction for both latencies and accuracies. The difference in accuracy may have occurred only in Experiment 3 for two reasons. The trend observed was consistent with the trend in Experiment 2, but Experiment 3 had greater power. Compared with Experiments 1a and 1b, thinking aloud in Experiment 3 may have forced participants to reason more explicitly, which may have yielded the predicted difference in accuracy.

Experiments 1a, 2, and 3 are consistent with the notion that reasoners rely on models to perform such inferences, but they do not indicate that they used mental models. Participants may instead have simply relied on external models, since they were always able to use paper and pencil in the experiments. However, the results of Experiment 1b suggest that individuals do also rely on mental models. The participants were not allowed to make any written workings, and yet single-clause problems again took less time than fully-relational problems.

Single-clause problems are easier than fully-relational problems. This result was predicted by our account, which integrates the model theory with the notion of effective relational complexity (see Goodwin & Johnson-Laird, 2005; Halford et al., 1998, 2005; Kroger, Holyoak, & Hummel, 2004). The result cannot be explained by the syntactic or formal complexity of individual premises, but it is open to alternative explanations. These explanations, however, do seem to need to take the complexity of the integration process into account and to allow that individuals tend to start reasoning from whichever premise, if any, that allows a ternary order to be constructed. That is, reasoners rapidly learn that they do not have to process premises in the order in which they are stated (see also Van der Henst et al., 2002).

The analyses of reasoners' strategies showed that they used a variety of tactics, which mainly concerned the construction of external models, and that they developed the three main a priori

strategies that such tactics make possible (see Experiments 2 and 3). None of these strategies relied on overt symbolic or formal manipulations. In the *ternary* strategy, they used one premise to construct a model of a ternary ordering and then used the other premise to place the fourth individual in the ordering, relying either on a binary relation or on the relation between the relations. In the *transitive* strategy, they used two clauses from the separate premises to construct a model of a transitive ordering of three individuals and one of the two premises to place the fourth individual in the ordering. In the *frame* strategy, they used one or both premises to construct a model of two individuals (i.e., the frame), next they used the premises to infer or to make a constrained guess about the order of the other individuals, and finally they checked the result against the premises. In principle, the ternary strategy is optimal, and the frame strategy is the most liable to yield errors. There were signs of such differences, but they were difficult to assess statistically because of the small numbers of individuals using the frame strategy. However, Experiment 3 did reveal that there was a reliable difference in latencies over the three strategies, with the ternary strategy the fastest and the frame strategy the slowest. Likewise, the ternary strategy transferred from one block of problems to another to a greater extent than did the other strategies.

The model theory's account of strategies postulates that individuals develop them "bottom up" as they try out different tactical steps, and that the resulting strategies constrain the sequences of tactical steps but do not freeze them into a deterministic sequence of operations (see Van der Henst et al., 2002). The experiments corroborated this account in four main ways.

First, in Experiment 2 the participants never made any remarks suggesting that they were constructing a strategy from the top down. Hence, they never said anything on the following lines: "With a problem of this sort, a sensible procedure is to work out an ordering of three individuals and then to search for a clause that allows the fourth individual to be integrated into this ordering." They showed no signs of identifying distinct

sorts of problem. However, the attempt to bias certain strategies in Experiment 3 appeared to have made reasoners more aware of the strategies that they were adopting. Several participants made metacognitive remarks, such as: "Now I'm going to try a new strategy." This greater metacognitive awareness may have contributed to the diminished strategic variation in Experiment 3.

Second, the protocols from both Experiments 2 and 3 suggested that the participants were trying out various tactical steps and assembling strategies from sequences of these steps. They were led from one tactic to another by the possibilities that the first tactic opened up rather than by a general plan dictating a sequence of tactical steps.

Third, different participants developed different strategies. They all tended to use more than one strategy, and in Experiment 2 they switched from one strategy to another reliably more often than was necessary. They also used more than one strategy in Experiment 3, but did not switch so often, presumably because the experimental conditions were effective in biasing the use of particular strategies.

Fourth, Experiment 3 successfully biased the particular strategies that individuals developed by manipulating the nature of the problem. When the premise affording a ternary relation occurred first, the participants tended to rely on the ternary strategy, which uses a premise to construct an initial ternary order. When this premise occurred second, and the two clauses yielding a transitive inference were printed in a different colour from the remaining text, then the participants tended instead to use the transitive or frame strategies. We did not anticipate the use of the frame strategy in this condition. However, an awareness of the transitive relation may have prompted some participants to select one of the end items in the ternary order to be an end element in a frame for the overall ordering. In this way, they bypassed the transitive strategy in favour of the frame strategy.

We conclude that individuals do not follow a deterministic procedure when they reason about the relational problems in our studies, but instead develop a variety of constraints on

sequencing their inferential tactics. These constraints yield strategies that are not deterministic, as in other sorts of reasoning (see also Bucciarelli & Johnson-Laird, 1999; Van der Henst et al., 2002). In contrast, when individuals carry out mental arithmetic, they appear to follow a deterministic procedure (see Johnson-Laird, 1993). This evidence, of course, does not rule out the possibility of "top-down" strategic processing, nor does it rule out the possibility of metacognitive insight developing with further experience with these problems. But, given only the experience that our experiments provided, individuals are unlikely to develop reasoning strategies in a self-conscious way.

In what ways might we have been misled by our results? It is conceivable that the need to think aloud led the participants to develop strategies quite different from those that they would develop when there was no requirement to think aloud. Yet, Experiments 1a and 1b, which had no requirement to think aloud, yielded the same overall pattern of latency results as did Experiments 2 and 3: Single-clause problems were faster than fully-relational problems. They did not however, yield quite the same pattern of accuracy results. We suspect that thinking aloud forced participants to reason more explicitly. We also suspect that it did not change their underlying strategies, although we cannot rule this possibility out. Hence, there is mixed support for the view that thinking aloud does not change the strategies that individuals develop (reasoning more explicitly does not require a change in strategy in our view), although other studies of reasoning have provided stronger evidence for this claim (e.g., Johnson-Laird & Hasson, 2003).

The model theory postulates that individuals construct mental models of the problems, and the protocols from the experiments were compatible with this account, showing that the participants tended to draw diagrams of the orderings. A sceptic might object that our problems unfairly bias individuals to use models, because the required answers were themselves diagrams. However, it seems unlikely that reasoners' strategies would have differed if we had required

them to write down a sentence as their final answer. The protocols reveal that reasoners used diagrams or models throughout the process of drawing an inference. They translated the input sentences into models and then manipulated these models in order to establish the correct ordering of individuals. The final model of the premises was not simply varnish on an altogether separate cognitive structure. It is conceivable that participants' use of diagrams do not reflect their underlying mental processes, in the same way that the use of words does not necessarily indicate the use of a verbal strategy in problem solving. Indeed, think-aloud methods reveal only the tactical steps that reasoners employ, and the different assemblages of these tactics give rise to different strategies. They do not yield access to lower level inferential mechanisms—that is, the processes that are responsible for building such diagrams or models (see, e.g., Van der Henst et al., 2002). Nevertheless, we take them to be strongly suggestive, if not decisive, evidence in favour of mental models.

The model theory postulates that comparative relations, such as those in our experiments, are represented on a single linear dimension. This system is capable of extending to other nonspatial comparatives, but it is necessary to keep track of the arguments of relations. For example, you can readily envisage the situation in which, say, Mark loves Anne more than Anne loves Jim. The model theory is also capable of extending to asymmetric relations, in which, say, Cordelia loves Lear more than Lear loves Cordelia, and to iterative relations, in which everyone loves Lear more than he loves anyone (see Cherubini & Johnson-Laird, 2004). However, none of these sorts of relation are within the immediate scope of our present account, and they would require additional representational machinery. Other higher order relations call for the introduction of separate dimensions, or for the computation of functions, for example:

John is taller than Mark by a greater percentage than Fred is heavier than Bob.

Relations such as these may lead to more difficult inferences, because they call for more complex models. What we have shown in other studies is that the higher the order of the relation that an inference depends on, the more difficult it is. Reasoning about relations between relations between relations is harder than the reasoning that we have investigated in the present study (see Goodwin & Johnson-Laird, 2005).

When individuals reason about relations between relations, they appear to rely on models of possibilities. The difficulty of an inference depends on the complexity of integrating the information from the two premises into a single model. These models may be internal (mental) or, alternatively, external diagrams—and our results suggest that they are at least sometimes mental. Yet, there may be other ways in which to reason about relations between relations, which our present data do not rule out. Reasoners do not rely on a single deterministic strategy, but instead spontaneously develop a variety of strategies for integrating information. The nature of the premises can bias the development of strategies. This phenomenon corroborates the principle in the model theory that individuals assemble inferential strategies bottom up from tactical steps rather than top down from some general plan.

Original manuscript received 3 June 2004

Accepted revision received 13 March 2005

PrEview proof published online 29 August 2005

REFERENCES

- Birney, D., & Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: An application of relational complexity to the knight-knave task. *Thinking and Reasoning*, *8*, 109–134.
- Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*, 247–303.
- Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language*, *28*, 564–575.

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

Cherubini, P., & Johnson-Laird, P. N. (2004). Does everyone love everyone? The psychology of iterative inferences. *Thinking & Reasoning*, 10, 31–53.

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

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, 54, 1–71.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, 112, 468–493.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, 16, 70–76.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, 21, 803–805.

Hartshorne, C., Weiss, P., & Burks, A. (Eds.). (1931–1958). *Collected papers of Charles Sanders Peirce* (Vols. 1–8). Cambridge, MA: Harvard University Press.

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

Huttenlocher, J., & Strauss, S. (1968). Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 7, 527–530.

Johnson-Laird, P. N. (1993). *The computer and the mind* (2nd ed.). London: Fontana Press.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory and Cognition*, 31, 1105–1113.

Kroger, J. K., Holyoak, D. J., & Hummel, J. E. (2004). Varieties of sameness: The impact of relational complexity on perceptual comparisons. *Cognitive Science*, 28, 335–358.

Kull, S. (2004, October 21). *The separate realities of Bush and Kerry supporters*. Retrieved March 5, 2005, from the Program on International Policy Attitudes Web site: http://www.pipa.org/OnlineReports/Pres_Election_04/Report10_21_04.pdf

Loft, S., Hill, A., Neal, A., Humphreys, M., & Yeo, G. (2004). ATC-lab: An air traffic control simulator for

the laboratory. *Behavior Research Methods, Instruments, & Computers*, 36, 331–338.

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

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

Van der Henst, J.-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science*, 26, 425–468.

APPENDIX A

The forms of the problems in Experiments 1A, 1B, and 2

| <i>Problem</i> | <i>Form</i> | <i>Correct response</i> |
|------------------|--|-------------------------|
| Single-clause | (A > B) > (C > D) (B > D) > (C > D) | A B C D |
| | (A > B) > (C > D) (C > A) > (C > D) | C D A B |
| | (A > B) > (C > D) (A > C) > (A > B) | A B C D |
| | (A > B) > (C > D) (D > B) > (A > B) | C D A B |
| Fully-relational | (A > B) > (C > D) (C > B) > (B > D) | A C B D |
| | (A > B) > (C > D) (C > A) > (A > D) | C A D B |
| | (A > B) > (C > D) (C > D) > (B > D) | A C B D |
| | (A > B) > (C > D) (C > D) > (C > A) | C A D B |

APPENDIX B

The forms of the problems used in Experiment 3

| <i>Block</i> | <i>Problem</i> | <i>Form</i> | <i>Correct response</i> | |
|------------------|----------------------------------|--|--|---------|
| Ternary strategy | Single-clause | (B > D) > (C > D) (A > B) > (C > D) | A B C D | |
| | | (C > A) > (C > D) (A > B) > (C > D) | C D A B | |
| | Fully-relational | (C > D) > (B > D) (A > B) > (C > D) | A C B D | |
| | | (C > D) > (C > A) (A > B) > (C > D) | C A D B | |
| | Transitive strategy ^a | Single-clause | (A > B) > (C > D) (B > D) > (C > D) | A B C D |
| | | | (A > B) > (C > D) (C > A) > (C > D) | C D A B |
| Fully-relational | | (A > B) > (C > D) (C > D) > (B > D) | A C B D | |
| | | (A > B) > (C > D) (C > D) > (C > A) | C A D B | |

^aThe clauses in bold were presented in red ink for this block.