Memory & Cognition
2002, 30 (1), 129-137

Generating alternatives: A key component in human reasoning?

STEPHEN E. NEWSTEAD
University of Plymouth, Plymouth, England

VALERIE A. THOMPSON
University of Saskatchewan, Saskatoon, Saskatchewan, Canada

and

SIMON J. HANDLEY
University of Plymouth, Plymouth, England

According to mental models theory, a key aspect of deductive reasoning is the production of alternative models that can falsify provisional conclusions. In the present paper, the possibility is investigated that there are individual differences in the ability to produce alternative models. The results indicate that some people do not proceed beyond the first model when they reason with syllogisms but that others do. Furthermore, the ability to generate alternatives can be independently measured by asking participants to generate different representations of pairs of premises. These findings support the predictions of mental models theory and also indicate the potential importance of alternatives generation as a measure of individual differences in processing style.

One of the most influential current theories in the field of human reasoning is that people construct mental models (Johnson-Laird & Byrne, 1991). According to this theory, when faced with a reasoning problem, people will first construct a representation of the premises using image-like tokens. They then see what conclusion (if any) follows from this and proceed to search for alternative representations (or models) in which that conclusion is not true. If no falsifying model can be found, the conclusion is accepted; otherwise, it is rejected, and the search for an alternative conclusion continues. The search for falsifying models lies at the heart of mental models theory.

Recently, a debate has been taking place as to whether people do spontaneously search for alternative models (Klauer, Musch, & Naumer, 2000; Polk & Newell, 1995). These theories are very similar in many respects to mental models but do not assume that falsification takes place. Instead, it is explicitly assumed that reasoners attempt to construct only a single model or representation of the premises and base their judgments of validity on that one representation. Indeed, Polk and Newell demonstrated that the inclusion of a falsification parameter did little to improve the fit of their models to the syllogistic reasoning data.

Empirical evidence is also beginning to accumulate that reasoners do not always search exhaustively for alternative models; indeed, there is recent evidence that, in syllogistic reasoning at least, people seldom go beyond the initial model that they construct (Evans, Handley, Harper, & Johnson-Laird, 1999; Newstead, Handley, & Buck, 1999). Evans et al. (1999) found that people tended to accept conclusions that were consistent with the preferred initial model that they constructed and to reject conclusions that were inconsistent. Errors occurred when the conclusion was consistent with the initial model but where there existed alternative, falsifying models. Participants seemed unwilling or unable to construct these alternative models. The Newstead et al. (1999) study involved participants' constructing diagrams of different colored shapes that were consistent with the premises of given syllogisms, under the assumption that these would reflect the models constructed. The results indicated that people tended to produce the same representation over and over again (though in different forms).

This is not to say that people are unable to search for alternative models: All the evidence suggests that they can. For example, if the conclusion is unbelievable, people are likely to try to construct an alternative model (Newstead, Pollard, Evans, & Allen, 1992). Search for alternatives can also be induced by the instructions given (Evans, Newstead, Allen, & Pollard, 1994), or by reducing the cognitive load (Roberts, 2000). Furthermore, the ease with which alterna-

This research was funded by an operating grant from the Natural Sciences and Engineering Research Council of Canada to V. A .T. The order of authorship for the first and second authors was determined randomly. Correspondence and reprint requests may be addressed to S. Newstead, Department of Psychology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, England (e-mail: snewstead@plymouth.ac.uk) or to V. Thompson, Department of Psychology, 9 Campus Drive, University of Saskatchewan, Saskatoon, SK, S7N 5A5 Canada (e-mail: valerie.thompson@usask.ca).

—Accepted by previous editorial team

tives can be brought to mind appears to be a critical factor. There is a wealth of evidence in conditional reasoning that the availability of alternative antecedents influences performance in causal reasoning (Cummins, 1995), the Wason selection task (Handley, Feeney, & Harper, in press), and everyday conditional reasoning (Byrne, 1989; Thompson, 1995, 2000). Similarly, the availability of alternative routes to the consequent also appears to be important (Cummins, 1995; Thompson, 1995, 2000).

There is thus ample evidence that task variables can affect the production of alternative models. In addition, there may also be individual differences in the willingness to search for alternatives, and the ability to construct alternative representations might be systematically linked to reasoning performance. This is a central theme of the present paper. The possibility of such individual differences has long been acknowledged by mental models theorists (e.g., Johnson-Laird, 1983) and has recently been demonstrated empirically by Bucciarelli and Johnson-Laird (1999). These authors used a task in which their participants were asked to dress stick figures. For example, they might be given the syllogism

Some of the chefs are musicians
All of the painters are musicians

and then were asked to dress the stick figures with chefs' hats, guitars, and artists' palettes in such a way that made the conclusion *Some of the chefs are painters* not possible. Their participants were able to do this, which once again illustrates that people are capable of falsifying conclusions if the task demands it. More important from the present perspective were the large individual differences in performance. Their best participant produced 95% correct responses, their worst just 25% (the participants were students at Princeton University).

Further evidence for the existence of individual differences in alternatives generation stems from the work of Markovits (e.g., 1984). He measured participants' ability to think of alternatives by simply asking them the following question: *When David has homework to do, he gets into a bad mood. I saw David after school today and he was in a bad mood. Can you imagine what would have put David in a bad mood*? Those participants who indicated possibilities other than homework were superior in conditional reasoning performance to those who mentioned just the one possibility.

A recent study by Torrens, Thompson, and Cramer (1999) produced a similar picture. In a study of conditional syllogistic reasoning, participants were first asked to draw as many diagrams as possible to represent the premises of two quantified syllogisms; the number of different diagrams produced was the *alternatives generation* score. This score correlated positively with logical performance on the conditional syllogisms and negatively with the effects of conclusion believability. In other words, the participants who produced more alternative representations were more likely to give logically correct answers and less likely to respond on the basis of belief. Furthermore, alternatives generation was the *only* measure that served to predict believability effects: A range of other measures, including need for cognition, open-mindedness, and intelligence, all failed to serve as predictors. Newstead, Handley, Harley, Wright, and Farrelly (2002) found that alternatives generation also predicted performance on abstract versions of the Wason selection task and on certain kinds of conditional reasoning tasks.

Our main hypothesis is that reasoners vary in their ability and inclination to construct alternative representations and that these individual differences are systematically related to performance on syllogistic reasoning tasks. The basic prediction is that people who generate more alternatives will be more likely to solve certain types of syllogisms correctly.

However, a correlation between the ability to search for alternatives and syllogistic reasoning is open to an alternative explanation, since both might be mediated by a third variable such as need for cognition (e.g., Klaczynski, Fauth, & Swanger, 1998) or general intelligence (e.g., Stanovich & West, 1998). One way to rule out this possibility is to include measures of these variables in the experimental design. By using this strategy, both Torrens et al. (1999) and Newstead et al. (2002) found that their measure of alternatives generation predicted deductive reasoning performance even after accounting for the effects of intelligence and the need for cognition.

An alternative approach, and the one adopted in the present study, is to use materials that on a priori grounds should discriminate between reasoners who search for alternatives and those who do not. Thus, we used problems whose solutions required a search for alternatives. On such problems, a correlation between the search for alternatives and reasoning performance would provide strong evidence that it is the search for alternatives itself and not a more global reasoning ability that is mediating performance.

In our experiment, we presented reasoners with pairs of premises and a conclusion to evaluate. The conclusion was either weak or strong (Evans et al., 1999), a distinction that is perhaps best illustrated by an example. The syllogism *No A are B*; *All B are C* is a multiple model syllogism, having, according to Johnson-Laird and Byrne (1991) the following models:

	(1)			(2)			(3)	
[a]			[a]			[a]		c
[a]			[a]		c	[a]		c
	[[b]	c]		[[b]	c]		[[b]	c]
	[[b]	c]		[[b]	c]		[[b]	c]

In this notation, square brackets indicate that the entity is exhaustively represented. Thus, in model (1), there are no *a*s that are either *b*s or *c*s; *b*s are exhaustively represented with respect to *c*s; and there are no other *c*s that are either *a*s or *b*s (though there may be other *c*s that are neither *a*s nor *b*s). Now, if this is the first and only model constructed, participants would probably indicate that the conclusion *No A are C* (or that *No C are A*) necessarily followed, and indeed these two conclusions are very common (see Evans et al., 1999). If, on the other hand, participants construct only model (2), they would reject these two conclusions.

A strong conclusion is one that is consistent with the first model constructed, whereas a weak conclusion is one that is not consistent with the initial model but is consistent with an alternative representation. We believe that many reasoners will accept a conclusion that is consistent with the initial model, but reject a conclusion that is inconsistent with this model (Polk & Newell, 1995). In other words, they will accept strong conclusions and reject weak conclusions. However, other reasoners will not be content with the outcome of this initial decision process and will search for alternative representations. Those who do so should be more likely, for example, to correctly reject invalid strong conclusions (i.e., conclusions that are consistent with the initial model but not with other possible representations). We hypothesize that individual differences in reasoning style will mediate whether or not a person will search for alternative representations and hence influence performance on specific types of conclusions. This is a very strong test of our hypothesis. In contrast, there are other cases in which the ability to construct alternative representations should not be correlated with performance since the correct answer will emerge from the initial model constructed (Handley, Dennis, Evans, & Capon, 2000).

A full set of predictions is presented in the introduction to Experiment 2. Experiment 1 was designed to determine which is the preferred initial model for the syllogisms we were using. These findings are interesting in their own right and are thus reported in some detail.

EXPERIMENT 1

In order to determine the initial model constructed, we presented the participants with syllogisms; in one condition, we asked them to draw diagrams to represent the relationships and, in the other, to give the first conclusion that came to mind. These tasks were both presumed to indicate the initial model constructed, and we were looking for multiple model syllogisms that produced the same initial model both across participants and across tasks.

In addition to providing information about the initial models, this study had a second aim in that it allowed us to examine the relationship between the two tasks. Since both were designed to determine the initial model constructed, and since mental models theory assumes that initial conclusions are derived from initial models, this theory would presumably predict that the two tasks should correlate.

Method

Participants. The 111 participants were introductory psychology students at the University of Saskatchewan who participated for course credit. The mean age of the students was approximately 19 years, and approximately 60% were female.

Materials. According to Johnson-Laird and Byrne (1991) there are 17 valid multiple model syllogisms, all of which were used in the present study. Two booklets were prepared containing these syllogisms in a single random order.

The first booklet contained instructions that asked the participants to draw the first diagram that came to mind when they combined the two premises. An example was given: *All of the jeans are blue and all of the jeans are denim* (a multiple model syllogism). The instructions explained that the task required them to draw a diagram showing how jeans, blue things, and denim are related to each other. A sample diagram was given in which the first premise was drawn as a circle representing jeans; this was contained within (i.e., was a subset of) a larger circle representing blue things. A further diagram indicated that the second premise could be combined with this diagram by drawing a circle representing denim things so that it contained the circle representing jeans and overlapped with the circle representing blue things. It was stressed that although there may be other ways of representing the premises, it was important to draw the first diagram that came to mind.

The instructions in the second booklet asked the participants to write down the first conclusion that came to mind. The same example premise pair was given, and the instructions asked the participants to think of a conclusion that linked the two nonrepeated items

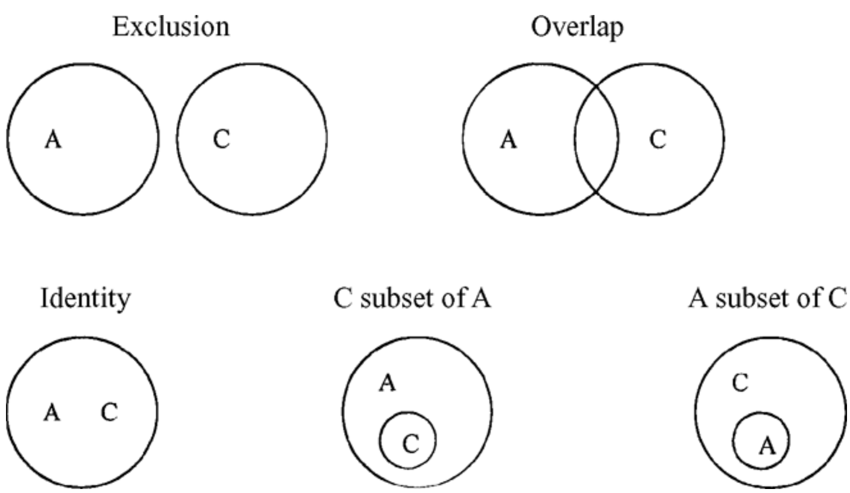

Figure 1. Possible Euler circle relationships between two sets A and C.

in the premises. The sample conclusion *Some of the blue things are denim* was given. It was explained that all answers needed to contain one of the standard quantifiers (*all*, *some*, *some. . . not*, and *no*) and that the important thing was to give the first conclusion that came to mind.

Each booklet contained the 17 pairs of premises in random order. The terms were all names of occupations or hobbies. Here is an example of one of the syllogisms used:

All of the piano players are doctors
and
None of the piano players are golfers.

Procedure. The participants were run in small groups, were randomly allocated to conditions (i.e., diagram or conclusion booklet), and were allowed as much time as they wanted to complete the task. There were 54 participants in the diagrams condition and 57 in the conclusion-production task.

Results and Discussion

On the diagrams task, the participants produced a wide variety of different diagrams that we inductively placed into 42 categories covering all the depicted relationships between the three sets (the size or shape of the diagrams used to depict the sets was deemed to be irrelevant). The 42 diagrams produced can be collapsed and classified solely according to the set relationship between the two end terms, A and C. The are just five possible relationships: identity, A as a subset of C, C as a subset of A, overlap, and exclusion (see Figure 1). The frequency with which these diagrams were given for each syllogism is presented in the left-hand part of Table 1. The conclusions given by the participants in the conclusion-production task are presented on the right-hand side of Table 1.

Although different participants were involved in the diagram and conclusion-production tasks, one would expect some relationship between the tasks. The diagrams produced in the diagram drawing task were assumed to indicate the first model produced by the participants and, as can be seen in Table 1, there was considerable consistency in the preferences shown. One would expect that the conclusions given in the second condition would reflect this, since they are presumably drawn from this first model. Because this is a between-subjects design, there is no way of examining this relationship directly, but it can be investigated indirectly. A fairly straightforward prediction is that the conclusion generated for a syllogism will, by and large, be consistent with the preferred model in the diagrams task. For those problems where the exclusion diagram is preferred, one would expect conclusions to be either *No A are C* or *No C are A*. For those syllogisms where the overlap relationship is preferred, one would expect *some* and *some. . . not* conclusions in either direction.

In order to examine this, the prediction was tested that the participants would tend to produce more conclusions that were consistent with the preferred model than would be expected by chance. For the exclusion relationship, which was the most common in the diagrams task, only two of the eight conclusions (those using *no*) were consistent with this; hence chance expectancy is .25. When the overlap relationship was preferred, half of the conclusions (those using *some* and *some. . . not*) were consistent with this; hence, the chance probability is .5. The preferred diagram and the conclusions consistent with it are shown in bold in Table 1. It can be seen that for 14 of the 17 syllogisms, conclusions consistent with the preferred model occurred more often than by chance, a difference that is significant on a sign test ($p < .01$).

Despite this general tendency, however, there are clearly some syllogisms where the conclusions were not consistent with the preferred diagram. The two most obvious ex-

Table 1
Diagrams and Conclusions Generated in Experiment 1

		Diagram					Conclusion							
	Syllogism	Exclusion	Overlap	Identity	AsubsetC	CsubsetA	Aac	Aca	Iac	Ica	Oac	Oca	Eac	Eca
1	Aab,Ocb	**73**	20	0	2	4	0	2	32	23	18	20*	**2**	**2**
2	Iab,Ebc	**68**	30	0	2	0	0	0	22	0	37*	0	**28**	**13**
3	Iab,Ecb	**76**	22	0	2	0	0	0	14	7	30*	2	**33**	**14**
4	Eab,Abc	**81**	17	0	0	2	0	0	10	2	4	2*	**60**	**21**
5	Eab,Ibc	**75**	21	0	0	4	2	2	11	7	9	7*	**46**	**17**
6	Eab,Icb	**73**	21	0	2	4	0	0	7	14	2	25*	**34**	**18**
7	Oab,Acb	**80**	18	0	0	2	0	4	44	11	22*	13	**2**	**2**
8	Aba,Abc	20	**58**	14	6	2	73	0	**22***	**2***	**0**	**0**	2	0
9	Aba,Ebc	**73**	20	0	6	0	6	0	6	0	9*	0	**68**	**11**
10	Aba,Obc	30	**59**	0	2	9	0	2	**30**	**9**	**54***	**2**	2	0
11	Aba,Ecb	**76**	24	0	0	0	2	4	7	2	7*	2	**43**	**33**
12	Iba,Ebc	**78**	18	0	4	0	0	0	15	2	31*	0	**35**	**17**
13	Iba,Ecb	**75**	21	0	4	0	2	0	11	2	21*	9	**13**	**43**
14	Eba,Abc	**76**	18	2	0	4	0	7	7	0	2	9*	**43**	**33**
15	Eba,Ibc	**72**	19	0	0	9	0	0	9	4	13	26*	**21**	**28**
16	Eba,Icb	**74**	23	0	0	2	2	0	9	6	4	36*	**28**	**15**
17	Oba,Abc	26	**67**	0	0	2	4	2	**27**	**13**	**6**	**42***	4	2

Notes—The logically correct conclusion is indicated by an asterisk. Entries in bold indicate the most popular diagram in the diagram construction task and the conclusions consistent with this in the conclusion task. A, *all*; E, *no*; I, *some*; O, *some. . . not*. Syllogistic figure is indicated by lower case letters, for example Aab,Ocb indicates All A are B, Some C are not B. The figures indicate the percentage of participants selecting each diagram or conclusion.

amples are Syllogisms 1 and 7 in Table 1. On both of these, the exclusion relationship was the preferred representation, but conclusions involving *no* were given by just 4% of the participants! The other major discrepancy is on Syllogism 8, where the preferred diagram was overlap but only 24% of the participants gave either *some* or *some. . . not* conclusions. The most popular conclusion was *All A are C*, given by 73% of people.

The degree of consistency between the diagram and conclusion-production tasks suggests that there are common processes mediating performance, and mental models provide a plausible account of what these processes might be. However, there were sufficient discrepancies to warrant caution in this conclusion. These discrepancies might well reflect task-specific demand characteristics. In the diagram task, for example, reasoners have a clear preference for exclusion relations, which might be because these are conceptually the easiest to represent diagrammatically. In contrast, a different set of biases, such as atmosphere bias, may operate in the conclusion-production task. It seems reasonable to conclude that performance on these tasks results from a combination of rational (i.e., model based) and heuristic processes.

Although the actual responses produced in this study are of interest in their own right, our main purpose in this study was to find syllogisms that consistently produced the same initial model with this population of participants. We selected eight such syllogisms in Experiment 2, the aim of which was to investigate differences in performance as a function of individual differences in alternatives generation.

EXPERIMENT 2

The central purpose of this experiment was to examine individual differences in performance on multiple model syllogisms that consistently produce the same initial model. The predictions are a little complex. In one condition, we used necessity instructions, under which only conclusions that necessarily follow should logically be endorsed. If the target conclusion is necessary (i.e., true in all models), one would expect it to be accepted by all participants; there should be no difference dependent on the search for alternative models since such a search is unnecessary. But if the conclusion is a strong one (i.e., true in the first model constructed but false in other models), one would expect such differences. The participants who construct only one model should accept strong conclusions as necessarily true, but those who construct alternative models should be more likely to find a disconfirming model. Hence, under necessity instructions, one would predict a correlation between alternatives generation and accuracy on strong conclusions.

In a second condition, we used possibility instructions, under which any conclusion that could be true should be accepted. If the conclusion is impossible, it should be rejected by all participants irrespective of whether they try to generate alternative models. However, things are different if the conclusion is a weak one (i.e., one that is not possible in the initial model but possible in other models that might be constructed). In this case, one would expect that those who generate alternatives should be more accurate than those who do not, since the former will be more likely to find the model in which the conclusion is possible.

In sum, we predicted that those who search for alternatives will show a different pattern of responding than those who do not. Under necessity instructions, those who search for alternatives should be more likely to reject strong conclusions; and under possibility instructions, they should be more likely to accept weak conclusions. Performance on necessary and impossible conclusions was not predicted to differ among reasoners.

Although the major focus was on the measure of alternatives generation used by Torrens et al. (1999), we also used other measures that may be related to the generation of alternatives. Many tests of creativity ask people to generate alternative responses, and we were interested to know whether one such test, the *uses of objects test*, correlated with either the alternatives generation task or syllogistic reasoning performance. In addition, we also explored the possibility of any correlations between these measures and a widely used measure of thinking style, the Rational Experiential Inventory (REI). This measure (and one of its precursors, need for cognition) has been claimed by some (e.g., Klaczynski et al., 1998) to be related to reasoning performance.

Method

Participants. There were 151 participants in this study, 64 males and 87 females, whose mean age was 19.4 years. They were all introductory psychology students at the University of Saskatchewan who took part for course credit.

Materials. We used the REI as described by Pacini and Epstein (1999). This is a 40-item self-report inventory that measures two main dimensions, rationality (the extent to which a person enjoys and engages in problem solving) and experientiality (the extent to which a person relies on intuition and past experience).

The uses of objects test was employed. This asked the participants to list as many uses as they could think of in 5 min for a brick and a paper clip (adapted from Hudson, 1966). The test was scored in terms of the total number of uses produced after the elimination of any obvious repetitions.

The alternatives generation task was modeled closely on that of Torrens et al. (1999). Four pairs of quantified premises were presented with neutral content. These were all multiple model syllogisms, though the content was different than that in the syllogistic reasoning task. The instructions were similar to those in the diagrams condition of Experiment 1. However, after the first overlap diagram had been presented, it was pointed out that other alternatives were possible, and a diagram presenting an overall subset relationship was given. The instructions indicated that the participants should draw as many diagrams as they could until they could think of no more. Their responses were scored in terms of the number of different but correct diagrams produced.

In the main syllogism task, we used eight syllogisms selected on the basis of the results of Experiment 1. We chose eight problems that reliably elicited the same first model as indicated by the findings in each of the two conditions. The problems used were Syllogisms 4, 5, 6, 9, 11, 12, 13, and 15 in Table 1. These were presented in a booklet in two groups of four. In one group of four, necessity instructions

Table 2
Conclusions Used in Experiment 2

Syllogism Number	Necessary	Strong	Weak	Impossible
4	Oca	Eac	Iac	Aca
5	Oca	Eac	Iac	Aca
6	Oca	Eac	Ica	Aca
9	Oac	Eac	Iac	Aac
11	Oac	Eac	Iac	Aac
12	Oac	Eac	Iac	Aac
13	Oac	Eac	Iac	Aac
15	Oca	Eac	Iac	Aca

Notes—A, *all*; E, *no*; I, *some*; O, *some. . . not.* Direction of conclusion (either A to C or C to A) is indicated in lower case letters.

were given indicating that a conclusion should be accepted only if it was necessarily true. The other set of four syllogisms used possibility instructions, explaining that a conclusion should be accepted if it was possible and rejected if it was impossible.

Each syllogism had four associated conclusions, one which was necessarily true, one which was strong (i.e., invalid but consistent with the initial model as determined in Experiment 1), one which was weak (i.e., possible but not consistent with the initial model), and one which was impossible. These conclusions are presented in Table 2.

The syllogisms in both conditions appeared as in the following example:

All the chefs are bankers
None of the chefs are workaholics
Therefore, some of the bankers are workaholics
Yes No

The task was to circle the correct answer (which varied, of course, with the instructions given). The content was chosen to be neutral, involving occupations, hobbies, and general descriptors. The eight syllogisms all had different content, and to avoid any materials effects, four different random combinations of material with syllogisms were constructed.

Procedure. The participants were run in a single large group. The order of presentation of the tasks was as follows: uses of objects, syllogisms, REI, and alternatives generation. The participants were presented with syllogisms under both necessity and possibility instructions, with the necessity condition being presented first to half of the participants and second to the other half. Four different syllogisms were presented in each condition. Under necessity instructions, two necessary and two strong conclusions were given, and under possibility instructions, two impossible and two weak conclusions were presented. The order in which the conclusion types were presented was counterbalanced across participants.

To summarize, each participant received eight syllogisms, four under necessity instructions and four under possibility instructions. Under necessity instructions, two syllogisms had conclusions that were necessarily true and two had conclusions that were invalid but consistent with the first model constructed (i.e., strong). Under possibility instructions, two problems had impossible conclusions and two had conclusions that were possible but not consistent with the first model constructed (i.e., weak). Each syllogism had just one conclusion to be evaluated, with the pairing of conclusion type with syllogism being systematically varied.

Results

Syllogisms task. Under the necessity instructions, the participants accepted more necessary (66%) than strong (52%) conclusions, a difference that is statistically significant [$t(152) = 2.88, p < .01$]. However, the proportion of strong conclusions accepted was very high and consistent with the suggestion that many participants did not look beyond the first model that came to mind. Under possibility instructions, more weak (40%) than impossible (10%) conclusions were accepted, and this difference was significant [$t(152) = 8.39, p < .001$]. Again, the relatively small number of weak conclusions that was accepted is consistent with the suggestion that some of the participants constructed a first model in which the conclusion was not possible and did not proceed to generate alternative models. Perhaps the most striking illustration of this is that reasoners were more likely to (incorrectly) accept a strong conclusion as necessarily following from the premises than they were to (correctly) accept a weak conclusion as possibly following from the premises [$t(152) = 2.48, p < .01$].

Individual differences measures. The correlations between the various measures are shown in Table 3. Of central interest to the present study are the correlations between the individual differences measures and performance on the syllogism task. The alternatives generation measure correlated negatively with acceptance of strong conclusions and positively with weak conclusions. This is exactly what was predicted and suggests that the participants who generated alternatives were more likely to search beyond their initial model. With strong conclusions, they were likely to find instances in which the conclusion was invalidated (i.e., was not necessary), whereas with weak conclusions, they found instances that showed the conclusion to be valid (i.e., where the conclusion was possible).

In contrast, alternatives generation was not expected to correlate with the rejection of impossible conclusions. As is indicated in Table 2, the correlation was close to zero,

Table 3
Correlations Between Measures Used in Experiment 2

	Uses of Objects	Rationality	Experientiality	Alternatives Generation	Strong	Impossible	Necessary	Weak
Uses of objects	–	.27*	.13	.17*	.12	−.08	−.01	.03
Rationality		–	.22*	.21*	.02	−.03	.02	.10
Experientiality			–	−.03	.11	.18*	.14	.07
Alternatives generation				–	−.17*	−.06	.21*	.20*
Strong					–	−.02	−.26**	−.26**
Impossible						–	−.18*	.04
Necessary							–	.14*
Weak								

Note—The measures of syllogism performance are based on the number of conclusions of each type accepted. $*p < .05$. $**p < .01$.

despite the fact that we had sufficient power to find a correlation (we had 80% power to detect a correlation of .2). Finally, we also observed that alternatives generation was positively correlated with correct acceptance of necessary conclusions.

The other individual differences measures were much weaker predictors of performance on syllogisms. None of them predicted performance on strong and weak conclusions: The only significant correlation was between experientiality and impossible conclusions. The two measures designed to test the ability to produce alternatives (uses of objects and the alternatives generation task) correlated with each other and with rationality. It would appear that there is some consistency among our various measures of the capacity to produce alternatives, suggesting that they reflect some common dimension, perhaps the ability to think flexibly. The fact that only the alternatives generation task predicted performance on syllogisms suggests that it may also measure a task-specific ability to generate alternative models in deductive reasoning tasks.

Correlations between measures on the syllogism task are also of interest. One would expect a negative correlation between acceptance of strong conclusions and weak ones, since the former suggests a failure to search for alternatives, whereas the latter indicates a tendency to conduct such a search. The predicted negative correlation can be seen in Table 2. Acceptance of necessary conclusions correlated with all of the other conclusions. The correlation was positive with the acceptance of weak conclusions and was negative with the acceptance of both strong and impossible conclusions.

Discussion

The present findings support the main predictions. There is evidence that individual differences in the production of alternative representations, as measured by the alternatives generation task, correlate in predictable ways with performance on weak and strong conclusions. In general, the more likely a person is to generate alternatives, the more likely they are to find the relevant model to disconfirm the conclusion suggested by the initial model. This clearly supports the claims made by the mental models theory that the search for falsifying conclusions underlies the reasoning process.

As predicted, alternatives generation did not correlate with performance on impossible conclusions. It did, however, correlate with necessary conclusions. There may be a very good reason for this. Necessary conclusions always used the quantifier *some not*, but we know from the results of Experiment 1 that the exclusion model is likely to be the first one constructed. Although logically *some not* is consistent with the exclusion relationship, we know that participants do not like to give this conclusion (Roberts, Newstead, & Griggs, 2001). As a consequence, reasoners may reject the conclusion unless they are able to construct an alternative representation that is consistent with a layperson's interpretation of *some not*. This means, of course, that people who can construct such alternative representations are more likely to accept necessary conclusions.

Our data are clearly consistent with the claim that people are capable of constructing alternative representations of syllogisms beyond the initial model constructed but do not always do so. Necessary conclusions were more commonly accepted than were conclusions that were consistent with the preferred model but not logically necessary. However, the latter were accepted by about half the participants, suggesting that they may not have gone beyond the first model. Similarly, impossible conclusions were rejected more often than were conclusions that were possible but not consistent with the initial model, but the latter were still rejected quite frequently. It would appear that on a significant number of trials, the participants constructed just a single model and did not attempt to falsify it.

A possible alternative explanation for our results is in terms of intellectual ability. Ability has been found to correlate with logical performance on syllogisms (e.g., Stanovich & West, 1998; Torrens et al., 1999) and it also correlates with alternatives generation (Newstead et al., 2002; Torrens et al., 1999). Perhaps, then, our results are little more than a demonstration of the importance of ability in syllogistic reasoning. There are, however, a number of reasons for believing that alternatives generation measures a process over and above general ability. Most importantly, we chose problems for which a correct answer could only have been reached by searching for alternatives. That is, the participants could only reject the strong conclusions and accept the weak conclusions if they searched beyond their initial representation of the premises. Thus, it seems unlikely that correlations with these measures can be attributed solely to a general logical ability instead of a more specific ability to construct alternative representations.

In addition, in our own data, we did not find a correlation between alternatives generation and performance on impossible conclusions as might have been expected if alternatives generation is closely related to intellectual ability. Furthermore, other studies have demonstrated that alternatives generation is in many ways a better predictor of some aspects of reasoning performance than are standard measures of intellectual ability (Newstead et al., 2002; Torrens et al., 1999). Finally, alternatives generation correlated with uses of objects (a measure of creativity) and rationality (a measure of cognitive motivation), neither of which are strong correlates of intellectual ability, and neither of which independently predicted syllogistic reasoning performance. This suggests that alternatives generation is more a measure of processing style than of intelligence. These arguments are not conclusive, but taken together, they do suggest that alternatives generation is not simply a measure of intellectual ability.

GENERAL DISCUSSION

The present data provide straightforward answers to the questions raised in the introduction. There is clear evidence for individual differences in the tendency to generate alternative representations of deductive reasoning problems. At one extreme, some people seem to produce just a single representation and to provide a conclusion consistent with this. They thus accept as necessary conclusions that are in fact compatible only with the initial model; and they also reject as not possible conclusions that are incompatible with the initial model but that are compatible with alternative models. At the other extreme, some people seem to attempt to falsify initial conclusions and do so by generating alternative models.

We have also shown that it is possible to measure independently this tendency to construct more than one model. The alternatives generation task is one that seems to provide a reasonable measure of this and that we now know to be a predictor of both syllogistic reasoning and various kinds of conditional reasoning (Newstead et al., 2002; Torrens et al., 1999). What is more, this measure seems to correlate with other, more generic measures of flexible thinking: the uses of objects test and rationality scores on the REI. This suggests that it is related to measures of flexible thinking.

In some ways, it may seem a little surprising that the alternatives generation measure proved to be a good predictor of syllogistic reasoning performance. The measure involves production of Euler circles to represent syllogisms, and this may lead to the adoption of different representational strategies to those used in syllogistic reasoning itself. Nevertheless, it is clearly a fairly demanding task, and the skills required to produce these diagrams seem to have some generality. The ability to generate alternatives may be a skill that more able thinkers develop, in the same way that experts develop skills that enable them to use long-term memory efficiently (Ericsson & Kintsch, 1995).

It is important to note that the success of the alternatives generation task in predicting performance does not commit us to a specific form of representation and, in particular, does not commit us to assuming that Euler circles are implicated in syllogistic reasoning. A number of theorists have claimed that reasoners use Euler circles as a means of representing syllogisms (Erickson, 1974; Ford, 1995; Stenning & Oberlander, 1995). We are using alternatives generation as a general measure of how well people can produce conceptually distinct representations; we do not commit ourselves to claims as to the form of those representations.

These findings have clear implications for mental models theory. On the one hand, the data support the work of those (e.g., Newstead et al., 1999; Polk & Newell, 1995) who have suggested that the search for alternative models is not a compulsory part of the reasoning process. Instead, the default procedure appears to be one in which reasoners proceed on the basis of a single model and are willing to accept a conclusion if it is consistent with that model. On the other hand, there is also clear evidence that reasoners are capable of constructing alternative models and do so when either individual inclinations or task-specific characteristics prompt them to. Moreover, the ability to construct alternatives is linked to logically correct reasoning performance.

Proponents of mental models theory will thus take considerable comfort from the present findings. The proposed centrality of alternatives generation to human reasoning is, of course, entirely consistent with this theory and indeed is derived from it. Furthermore, the similarities between diagram generation and initial conclusions observed in Experiment 1 are exactly what would have been predicted by this theory (though the discrepancies on some syllogisms may also be cause for concern).

Explanations of our findings in terms of other theories are possible but seem less natural. For example, mental logicians such as Rips (1994) could no doubt argue that alternatives generation is related to some more general logical ability and that the correlations between this measure and reasoning performance are mediated by this more general ability. However, although alternatives generation correlates with intellectual ability, we have argued that it is more a measure of processing style than of reasoning ability. There is nothing in the mental logic approach that captures why alternatives generation should be a key component in human reasoning. In contrast, generating alternatives is at the heart of mental models theory, and this view provides a natural and compelling account of the findings presented here.

REFERENCES

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

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, **31**, 61-83.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, **23**, 646-658.

Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. L. Solso (Ed.), *Theories in cognitive psychology: Loyola symposium* (pp. 305-330). Potomac, MD: Erlbaum.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245.

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

Evans, J. St. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, **6**, 263-285.

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

Handley, S. J., Dennis, I., Evans, J. St. B. T., & Capon, A. (2000). Individual differences and the search for counter-examples in syllogistic reasoning. In W. Schaeken, A. Vandierendonck, & G. de Vooght (Eds.), *Deductive reasoning and strategies* (pp. 241-265). Mahwah, NJ: Erlbaum.

Handley, S. J., Feeney, A., & Harper, C. (in press). Alternative antecedents, probabilities, and the suppression of inferences in Wason's selection task. *Quarterly Journal of Experimental Psychology*.

Hudson, L. (1966). *Contrary imaginations: A psychological study of the English schoolboy*. Harmondsworth, U.K.: Penguin.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.

Klaczynski, P. A., Fauth, J. M., & Swanger, A. (1998). Adolescent identity: Rational vs. experiential processing, formal operations, and critical thinking beliefs. *Journal of Youth & Adolescence*, **27**, 185-207.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, **107**, 852-884.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology*, **75**, 367-376.

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition*, **27**, 344-354.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2002). *Individual differences in deductive reasoning.* Manuscript submitted for publication.

Newstead, S. E., Pollard, P., Evans, J. St. B. T., & Allen, J. L. (1992). The source of belief bias in syllogistic reasoning. *Cognition*, **45**, 257-284.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality & Social Psychology*, **76**, 972-987.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, **102**, 533-566.

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

Roberts, M. J. (2000). Strategies in relational inference. *Thinking & Reasoning*, **6**, 1-26.

Roberts, M. J., Newstead, S. E., & Griggs, R. A. (2001). Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning*, **7**, 173-204.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, **127**, 161-188.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, **19**, 97-140.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology*, **49**, 1-58.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, **76**, 209-268.

Torrens, D., Thompson, V. A., & Cramer, K. M. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking & Reasoning*, **5**, 1-28.

(Manuscript received October 23, 2000;
revision accepted for publication August 24, 2001.)