THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2004, *57A* (?), 000–000

Working memory and syllogistic reasoning

David E. Copeland and Gabriel A. Radvansky
University of Notre Dame, Indiana, USA

The purpose of this study was to examine the relation between working memory span and syllogistic reasoning performance. In addition, performance for the reasoning task was compared to predictions made by mental model theory and the probability heuristics model. According to mental model theory, syllogisms that require the use of more mental models are more difficult. According to the probability heuristics model difficulty is related to the number of probabilistic heuristics that must be applied, or (for invalid syllogisms) inconsistencies between the derived and correct conclusion. The predictions of these theories were examined across two experiments. In general, people with larger working memory capacities reasoned better. Also, the responses made by people with larger capacities were more likely to correspond to the predictions made by both mental model theory and the probability heuristics model. Relations between working memory span and performance were also consistent with both theories.

Stanovich and West (2000) have argued that differences between normative and descriptive performance in reasoning is due to four factors: performance errors, computational limitations, applying the wrong normative model, and alternative task construals. This paper focuses on the first and second factors: performance errors and computational limitations. Computational limitations are addressed in terms of working memory capacity. Simply, the larger a person's working memory capacity, the better he or she should be able to reason. Performance errors are viewed as a result of computational limitations. That is, people who have smaller working memory capacities may opt to use simpler strategies to reduce the demands on working memory. This paper assesses the relation between syllogistic reasoning and working memory capacity, including an examination of whether capacity affects strategy use. Working memory capacity was measured using the operation span test (Turner & Engle, 1989). This test assesses how many items a person can temporarily hold in memory while he or she processes additional information (i.e., a maths problem).

Correspondence should be addressed to David Copeland, 118 Haggar Hall, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA. Email: copeland.9@nd.edu

We would like to thank Steve Boker, Angelina Copeland, Ken Gilhooly, Mike Oaksford, Michael Wenger, and an anonymous reviewer for their comments on earlier drafts of this manuscript. We would also like to thank Diane Berish, Jackie Curiel, Lacey Nielson, Kris White, and Sarah Wieber for their assistance in collecting the data. This research was supported in part by a grant from the Army Research Institute, ARMY-DASW01-99-K-0001.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980343000846

Categorical logic

Categorical logic, a type of deductive logic, involves drawing conclusions based on sets of premises that are assumed to be true. As long as the premises are valid, the conclusions appropriately derived from them are also valid. In categorical logic, concepts are expressions of class membership and are presented in the form of a *syllogism*. In a syllogism, the first premise describes the relation of a subject and a middle term (e.g., All A are B), and the second describes the relation of the middle term with a predicate (e.g., All B are C). Based on this, one has to draw a conclusion about the subject and predicate (e.g., Therefore, all A are C) or to determine that there is no valid conclusion.

There are four forms of categorical logic, each with a different quantifier:

A. All A are B (universal affirmative)
E. No A are B (universal negative)
I. Some A are B (particular affirmative)
O. Some A are not B (particular negative)

In addition, there are four ways, known as "figures", in which the subject, middle term, and predicate can be arranged:

1. AB-BC
2. BA-CB
3. AB-CB
4. BA-BC

The first two figures are *asymmetrical* because the middle term is in different places in the two premises. The last two are *symmetrical* because the middle term is in the same place for both premises. The combination of these forms and figures yields a total of 64 possible syllogisms.

The experiments here focus on categorical logic because it is a simple type of deductive logic, with a small number of possible syllogisms. This makes the assessment of the relation between logical reasoning and working memory easier to track. In addition, there has been an extensive amount of research on how people solve these syllogisms, and a number of reliable effects have been described.

Mental model theory

One of the most prominent theories of human syllogistic reasoning is the mental model theory developed by Johnson-Laird (Bara, Bucciarelli, & Johnson-Laird, 1995; Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). The idea is that people draw conclusions using mental models. Mental models are mental representations based on the state of affairs in the world. People construct them based on the terms and quantifiers in the premises. These mental representations simulate the possible ways that the terms can relate to one another. The mental models that are constructed can be compared to one another to contrast different situations that might be possible based on the premises.

According to mental model theory, people go through three stages. During the first stage, a model is constructed for the information stated in the first premise. Next, the information from the second premise is added to the model. In the second stage, people use this model to draw a conclusion. Finally, in the third stage, people test this initial conclusion by constructing alternative models. These alternative models are then compared to the conclusion to determine whether it is valid. If it is not, then another conclusion is drawn that is consistent with the alternative models. If a conclusion is true for all of the models, then that conclusion is the response. If, however, no conclusion is consistent with all of the models, then people respond, "no valid conclusion".

According to mental model theory, syllogisms that require the construction of fewer models are easier than those that require the construction of more models. More errors result for multiple model syllogisms when an individual incorrectly selects a conclusion from one model without considering the others. This could be due in part to an inability to successfully coordinate multiple models in working memory.

An example of an easy one-model syllogism is,

All A are B
All B are C

The first premise, "All A are B", can be represented like this:

[A] B
[A] B

Each line of the model represents an occurrence of a term, with an arbitrary number of occurrences. The brackets indicate that a term is exhaustively represented. As can be seen, the premise "All A are B" is represented by placing a "B" next to every "A". As noted earlier, the converse cannot be assumed to be true. Thus, the "B" terms are not surrounded by brackets, representing the possibility that there is a "B" that is not an "A".

The second premise, "All B are C", can be easily added to the model. Whenever there is a "B", there must be a "C":

[[A] B] C
[[A] B] C

The "B" terms all have brackets indicating that every "B" is a "C", and the "C" terms do not have brackets because there might be a "C" that is not a "B". The "A" terms have double brackets because every "A" is a "B", and because every "B" is a "C", leading to the fact that every "A" is also a "C". Hence, the correct conclusion can be drawn: "Therefore, all A are C."

Multiple model syllogisms are more difficult. For example, three models can be constructed for the syllogism,

All B are A
No B are C

These are:

1. [A [B]]
   [C]

2. [A [B]]
   A [C]
   [C]

3. [A [B]]
   A [C]

Based only on the first model, one might incorrectly conclude “No A are C” or “No C are A”. Using the first and second models, another incorrect conclusion, “Some C are not A” might be drawn. Finally, when all three models are considered, there is one conclusion that is supported: “Some A are not C.”

Mental model theory also explains why people might respond that there is no valid conclusion. An example of a multiple model syllogism that does not have a valid conclusion is,

No A are B
No B are C

Here, two models could be constructed:

1. [A]
   [B]
   [C]

2. [A] [C]
   [B]

Based upon the first model, one could conclude that “No A are C”. However, the second model supports the conclusion, “All A are C”. These conclusions conflict with one another, so by using both models, it can be determined that there is no valid conclusion. Thus, if multiple models are constructed people can determine that they are in conflict and that there is no valid conclusion.

Probability heuristics model

Another theory of how people solve syllogisms is the probability heuristics model (Chater & Oaksford, 1999). This model considers reasoning to be a combination of heuristics and probabilistic inferences. According to this model, people first use the *min heuristic* to arrive at a conclusion quickly. Following this, people can draw a probabilistic inference from it. For example, if the initial conclusion was “All A are C”, it is probable that “Some A are C” is true as well. The third step is to use a heuristic to determine the order of the terms.

After reaching a conclusion, people test, or reconsider, it. First, people vary in how confident they are in proportion to the informativeness of the premises. People should be more confident if a premise contains the quantifier “All” than if “Some . . . not” is in the premises. Second, if the conclusion that is reached contains the quantifier “Some . . . not”, then people should not be confident. If people are not confident, they respond with “no valid conclusion”.

Using these steps, the probability heuristics model predicts three conclusions for each syllogism. The first is based on the min heuristic and determining the order of the terms. The second is based on the probabilistic inference. Finally, the third predicted conclusion is “no valid conclusion”, based on a person’s confidence when testing the conclusion.

For valid syllogisms, predictions made by the probability heuristics model are similar to those made by mental model theory. Syllogisms that require only two steps (e.g., min and attachment heuristics) correspond to one-model syllogisms. Syllogisms that require three steps correspond to two-model syllogisms, and syllogisms that require four steps correspond to three-model syllogisms. Thus, the probability heuristics model also predicts that one-model syllogisms are easier than valid two-model syllogisms, which are easier than valid three-model syllogisms.

Working memory capacity

Working memory is an important component in the active processing of information. It is the aspect of cognition where information is actively maintained, manipulated, and elaborated. Some researchers have defined working memory in terms of a single system (e.g., Cowan, 1995), where it is the portion of long-term memory that is activated above a certain threshold. Other researchers have described working memory as a multiple component system, which can consist of structures such as a central executive and slave systems (e.g., Baddeley & Hitch, 1974). While there are differences in opinion as to how working memory is structured, most researchers agree that it is the type of memory associated with the active manipulation as well as the temporary retention of information.

Many researchers have operationalized working memory ability in terms of capacity or span (see Daneman & Merikle, 1996). It is measured by using a span test, such as the operation span test (Turner & Engle, 1989), in which people must keep track of a set of information (e.g., words) while they process another set of information (e.g., verify a maths problem). There are different interpretations of what span measures. One idea is that it measures the amount of resources a person has available (Baddeley, 1986). A second is that it measures how well a person can use controlled attention (Kane, Bleckley, Conway, & Engle, 2001). Alternatively, there is the idea that it measures how well people are able to deal with interference (Lustig, May, & Hasher, 2001). We make no distinctions between these views because they are all consistent with the idea that someone who has a large span is better at coordinating information in working memory than someone with a smaller span.

Logical reasoning and working memory

Working memory capacity should be important to logical reasoning. The mental model account of syllogistic reasoning is that "the effects of both number of models and figure arise from an inevitable bottleneck in the inferential machinery: the processing capacity of working memory, which must hold one representation in a store, while at the same time the relevant information from the current premise is substituted in it . . . the integration of premises has to occur in working memory, unless the subjects are allowed to use paper and pencil so as to externalize the process" (Johnson-Laird, 1983, p. 115). Thus, one would expect that people with larger working memory capacities to be better at syllogistic reasoning.

The idea that logical reasoning involves working memory has been explored to some degree. Gilhooly and colleagues (Gilhooly, Logie, Wetherick, & Wynn, 1993; Gilhooly, Logie, & Wynn, 2003) have done a number of dual-task studies. In these studies, people were given the task of drawing a conclusion to a syllogism. In addition to this, people also had to perform an auxiliary, or distracter, task at the same time. These auxiliary tasks included articulatory suppression, concurrent tapping, and random number generation. The results showed that reasoning and random number generation interfered with one another, and that reasoning also interfered with articulatory suppression. When the premises are presented sequentially (instead of simultaneously), reasoning also interfered with concurrent tapping. Gilhooly (1998) suggested that syllogistic reasoning uses all aspects of Baddeley and Hitch's (1974) model of working memory: the central executive and both slave systems. However, the visuo-spatial scratchpad played a smaller role, and its use was only observed when the premises were presented sequentially.

In general, these studies demonstrate that reasoning uses working memory resources. The question addressed here is, do working memory span tests predict syllogistic reasoning performance? Based on mental model theory, the more models that are needed to draw a correct conclusion, the more working memory resources should be taxed. People with higher working memory spans should perform better on syllogisms that require multiple models.

There are very few studies that have examined the relation between working memory span and categorical reasoning. Those that have been done have used measures of working memory span that have been criticized (i.e., digit span; Bara et al., 1995), have used working memory span tests that were similar to the reasoning tasks used (Kyllonen & Christal, 1990), or have focused on effects of ageing (Gilinsky & Judd, 1994). The current study avoids these shortcomings by using a more accepted measure of working memory span—the operation span test (Turner & Engle, 1989)—and by focusing on the relationship between span and reasoning for young adults (i.e., college undergraduates).

EXPERIMENT 1

The aim of Experiment 1 was to examine the relation between working memory span and syllogistic reasoning. According to mental model theory, models must be constructed and coordinated to arrive at a correct conclusion. It seems reasonable that these processes use working memory resources. However, it is possible that some people do not construct models, but base their conclusions on the "mood" of the premises (Gilhooly et al., 1993; Woodworth & Sells, 1935). Such a process would not tax working memory as much as constructing models. Under such circumstances, it would be expected that there would be little or no relationship between span and syllogistic reasoning performance.

While these strategies have not received support in previous research (see Evans, Newstead, & Byrne, 1993), they are considered here because it is possible that people who differ in working memory capacity also differ in the strategies that they use. For example, people with a large span might be better able to construct multiple models and manipulate the premises. However, people with a small span might find this more difficult and adopt simpler heuristic strategies. Some support for this idea has been found under dual-task conditions (Gilhooly, Logie, & Wynn, 1999). Secondary tasks were more disruptive for people who did better on an initial syllogistic reasoning task. That is, those people who performed poorly in the initial task were less affected by the introduction of a secondary task. Baddeley and Logie (1999) compared this finding to one by Rosen and Engle (1997), where individuals with high span scores showed a larger decrement as a result of a secondary task than did individuals with low span scores. These results suggest the possibility that people with a larger working memory span were using a more complicated strategy than people with a smaller span. The rationale for this argument is that a secondary task is more likely to disrupt more complex processing than simpler processing. If the people with a smaller span were using a complex strategy (e.g., mental models), then their performance should have suffered with the addition of the secondary task. However, performance was much more disruptive for people with a larger span.

Working memory span

The current study used the operation span test (Turner & Engle, 1989). This is a complex span test because not only does it include the temporary storage of information (i.e., retain a series of words), but it also includes processing of information (i.e., verifying a maths problem) as well. This is important because working memory is typically defined as consisting of both the short term retention and manipulation of information (e.g., Baddeley & Hitch, 1974). The operation span test has been used to study the relation between span and many different cognitive abilities (e.g., Conway, Tuholski, Shisler, & Engle, 1999) and is thought to measure general attentional control (Kane et al., 2001).

Strategies and predictions

The strategies that were tested in Experiment 1 were the use of mental models, probabilistic heuristics, or simple heuristics such as atmosphere and matching. Predictions for mental model theory were taken from the Syllog computer program developed by Johnson-Laird (1992). For each syllogism, this program constructs representations of mental models. It also provides conclusions, in both directions (i.e., A–C and C–A), based on these models. Predictions were based on conclusions that would be drawn if people only constructed one model, if they constructed two models (if applicable), and if they constructed three models (if applicable). Thus, the number of predicted responses for each syllogism ranged from two, for one-model syllogisms, to five, for three-model syllogisms.

The probability heuristics model makes three predictions for each syllogism. The first was derived by applying only the min and attachment heuristics. Next, the p-entailment heuristic was used to determine the second prediction. Finally, the last prediction for each syllogism was "No valid conclusion", based on the testing heuristics (i.e., lack of confidence in the response).

The predictions for atmosphere and matching were created by following the rules of each heuristic. This led to two predictions for each syllogism, one in each direction, for both heuristics. Atmosphere states that if at least one quantifier is particular (i.e., "some" or "some . . . not"), the conclusion has a particular quantifier. A second rule states that if at least one quantifier is negative (i.e., "no" or "some . . . not"), the conclusion has a negative quantifier. According to matching, people use the most conservative quantifier from the premises in their conclusion.

To illustrate these predictions, consider the example syllogism below,

No A are B
All B are C

According to Syllog, people might construct one or more of the following models:

1.	[A]			2.	[A]		C	3.	[A]		C
	[A]				[A]				[A]		C
		[B]	C			[B]	C			[B]	C
		[B]	C			[B]	C			[B]	C

Using only the first model people might respond with either “No A are C” or “No C are A”. If the second model was considered, the conclusions would be “Some A are not C” or “Some C are not A”. Finally, if the third model was also constructed, people would be likely to respond that there is no valid conclusion. Thus, there are five predictions made by mental model theory for this syllogism.

For the probability heuristics model, applying the min heuristic leads to the selection of the quantifier “No”. Then, the attachment heuristic is used to determine the order of the terms, “A–C”. Thus, the prediction is “No A are C”. When using the p-entailment heuristic people make a probabilistic inference, leading to the predicted conclusion “Some A are not C”. Finally, if the person is not confident in their response, they will state that there is no valid conclusion. In total, the probability heuristics model makes three predictions.

In terms of the simple heuristics, both would make two predictions: “No A are C” and “No C are A”. According to atmosphere, both quantifiers are universal, and one is negative, so the conclusion should use “No”. For matching, “No” is more conservative than “All”, so it also predicts that “No” should be used. As a reminder, these heuristics do not predict a specific order of the terms.

As can be seen, there are differences in the number of predictions that are made. Mental model theory and the probability heuristics model typically make more predictions. For example, there are syllogisms for which atmosphere makes only two predictions while mental model theory makes five. This discrepancy gives the second strategy an advantage in predicting responses. To address this problem, weighted scores were analysed. This was done by dividing the number of predicted responses by the total number of possible responses (i.e., nine; two with each of the four quantifiers plus “No valid conclusion”) and then subtracting that value from one. This calculation was done for each syllogism where the conclusion was correctly predicted. If a response was not predicted, then that strategy received a score of zero for that syllogism. Finally, the scores for each syllogism were summed for a total weighted score. This was done for each strategy. Thus, if a strategy predicts more possible responses for a syllogism, that strategy has a lower potential weighted score than has a strategy that predicts fewer responses. In other words, the more predictions that are made, the greater the penalty.

To illustrate this analysis, consider the syllogism “No A are B, All B are C”. If a person were to respond with the conclusion, “No A are C”, a weighted score would be calculated for each strategy. Both of the simple heuristics make two predictions, so they would get a score of 7/9, or 0.778. This is the number of predictions (i.e., two) divided by the total number of responses (i.e., nine), which is then subtracted from one. For the probability heuristics model, the score would be 6/9, or 0.667, because it makes three predictions. Finally, mental model theory would have a score of 4/9, or 0.444, because it makes five predictions. If, instead, a person responded with, “Some A are not C”, then atmosphere and matching would have scores of zero because they do not predict that response. Mental model theory and the probability heuristics model would have the same scores that were just calculated, 4/9 and 6/9, respectively. This weighted scoring was done for all syllogisms, and each of these scores were summed to get a total weighted score for each strategy.

Another important point is that for every syllogism, the predictions made by the simple heuristics (atmosphere and matching) are a subset of the predictions made by mental model theory. Specifically, these predictions almost always (except when the two heuristics have different predictions) are the same as the predictions of mental model theory for the construction of only a single model. Thus, for these responses, it is unclear whether people are using a simple heuristic or constructing a single model.

In terms of working memory, if people are using mental models or probabilistic heuristics, then people should be more accurate for syllogisms that require the use of fewer models or heuristics. In addition, those individuals that have larger working memory spans should be more likely to be accurate for syllogisms that require the use of more models or heuristics. Thus, working memory span should be related to syllogistic reasoning, especially for those syllogisms that require the use of multiple models or heuristics.

People that are using a complex strategy rather than simple heuristics will be more likely to respond with "no valid conclusion". The simple heuristics never predict this response. While there is overlap here for predictions made by mental model theory and the probability heuristics model, there are differences in their predictions of responses, as well as the number of responses predicted per syllogism.

Method

Participants

A total of 159 people, 95 females and 64 males, participated. They consisted of 98 undergraduates from the University of Notre Dame and 61 undergraduates from Indiana University South Bend who participated in exchange for payment or partial course credit. Data from three additional participants were eliminated because they did not complete all of the tasks. All participants were native English speakers.

Materials and procedure

All of the tasks were administered on desktop or laptop PC computers.

Operation span (Turner & Engle, 1989). People had to read aloud and verify the solution to a two-step maths problem. The first step involved multiplication or division, and the second step involved addition or subtraction—for example $(2 \times 3) - 2 = 4$. People verbally indicated whether the answer was correct, and the experimenter pressed a key to advance. Following this, a word was presented for one second. People were instructed to read this word aloud. The items were presented in a random order for each participant, and the problem—word pairings were also randomized. After completing a set of problems, people recalled the words verbally, and the experimenter typed the words. Following a method used by LaPointe and Engle (1990), a score was determined by summing the total number of items in correctly recalled sets. For example, if a person got three sets of two correct, plus two sets of three, then the score would be 12.

Syllogistic reasoning. All 64 possible syllogisms were used. The subject, reference, and middle terms in each syllogism were based on occupations, hobbies, or interests (e.g., surgeons, cyclists, coffee drinkers, etc.). These terms were randomly assigned to each syllogism for each participant. An example of a possible syllogism is:

All cyclists are coffee drinkers
All coffee drinkers are surgeons

Hobbies/interests and occupations were used to ensure that each of the situations that were described would be plausible to the reader and to minimize the belief bias effect.

The premises of each syllogism were presented on the computer screen along with nine possible conclusions. Eight of the choices were created by making two choices of each form (i.e., A, E, I, and O): one with the subject term first and one with the reference term first (e.g., All cyclists are surgeons, All surgeons are cyclists). The ninth choice was "No valid conclusion". The premises and the choices were presented simultaneously, and people had an unlimited amount of time to select their response. The choices were always presented in the same order.

Under certain circumstances, it is possible that a person could be unsure as to which conclusion should be selected (e.g., if "Some A are C" is true, "Some C are A" is also true). People were instructed that if they thought that more than one conclusion was true, they should respond with the one they reached first.

Experimental context

These measures were embedded in the context of a larger study.[1] They were administered in this order: (a) vocabulary, (b) speed, (c) situation identification, (d) word span, (e) reading span, (f) updating task, (g) spatial span, (h) text comprehension and recognition, (i) operation span, (j) logical reasoning, and (k) fan effect. Testing was done either across two 1.5-hour or three 1-hour sessions. If done over 2 days, Tasks a–h were done on Day 1, and tasks i–k were done on Day 2. If done over 3 days, Tasks a–f, g–i, and j–k were done on Days 1, 2, and 3, respectively. The results of these tasks are considered elsewhere (e.g., Radvansky & Copeland, 2001).

Results

All *p* values are less than .05 unless reported otherwise. For correlation measures, a bootstrapping method (Efron & Tibshirani, 1998) with 1,000 bootstrap samples was used to correct for any influence of skewed distributions. Confidence intervals of 95% are reported for these analyses.

Accuracy

Overall, participants' accuracy ranged from 14% to 89% on the syllogistic reasoning task. The number of models was a significant factor, $F(2, 316) = 961.10$, $MSE = 0.015$ (see Table 1). Performance for one-model syllogisms was better than that for two-model syllogisms, $F(1, 158) = 975.87$, $MSE = 0.019$, which was better than that for three-model syllogisms, $F(1, 158) = 48.99$, $MSE = 0.010$. These results are consistent with mental model theory.

[1]Additional analyses were done using the word, reading (Daneman & Carpenter, 1980), and spatial (Shah & Miyake, 1996) span tests. These span tests showed very similar patterns to the operation span test.

TABLE 1
Accuracy[a], response times[b], and confidence ratings[c] for number of models in Experiments 1 and 2

	Experiment 1		*Experiment 2*					
	Accuracy		*Accuracy*		*RT*		*Confidence*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overall	43	10	45	17	30	11	4.9	1.1
One-model	86	14	87	16	25	9	5.6	1.1
Two-model	39	20	40	21	29	12	4.7	1.2
Three-model	31	17	34	20	33	12	4.8	1.2

[a] In percentages. [b] In s. [c] Out of 7.

Strategy use

Weighted scores. Mental model theory has an advantage in our strategy use comparison because it predicts more responses than do the simple heuristics. To correct for this, weighted scores (see Table 2) were computed for the strategies for each person to determine how well each predicted responses. Each syllogism received a score, using the formula, 1– (# of predicted conclusions / 9). These scores were summed to create a weighted score for each strategy.

The weighted score for mental models was only slightly, but significantly, larger than the weighted score for the probability heuristics model, $F(1, 158) = 4.18$, $MSE = 6.065$. Both mental model theory and the probability heuristics model had weighted scores that were larger than atmosphere, $F(1, 158) = 144.73$, $MSE = 18.248$, and $F(1, 158) = 70.22$, $MSE = 30.605$, respectively. Finally, atmosphere had a larger weighted score than matching, $F(1, 158) = 44.30$, $MSE = 4.623$. Even though mental model theory predicts a wider variety of responses it still does the best job, although only slightly better than the probability heuristics model, of predicting performance when this inherent advantage is taken into account.

TABLE 2
Percentage of responses that corresponded to strategies, and weighted scores in Experiments 1 and 2

	Experiment 1				*Experiment 2*			
	% resp.		*Weighted score*		*% resp.*		*Weighted score*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mental models	85	12	32.9	4.5	84	12	32.8	4.8
PHM	76	12	32.3	5.2	76	13	32.6	5.3
Atmosphere	54	12	27.1	6.2	52	12	25.8	6.1
Matching	51	12	25.5	6.1	49	13	24.2	6.3

Working memory and syllogistic reasoning

Overall, the correlation between working memory span and syllogistic reasoning was significant, r = .422,Conf.(.271 $\leq \rho_I \leq$ .559) = .95. The larger one's span the better one's reasoning performance. Working memory was correlated with performance for one-, two-, and three-model syllogisms; r = .185, Conf.(.029 $\leq \rho_I \leq$ .315) = .95; r = .399, Conf.(.248 $\leq \rho_I \leq$ .357) = .95; and r = .370, Conf.(.223 $\leq \rho_I \leq$ .512) = .95, respectively. While these were all significant, the correlations were stronger for multiple model syllogisms. This suggests that a larger span is more helpful for reasoning when more models or probabilistic heuristics are involved.

Working memory and strategy use

Span was related to the predictions made by mental model theory, r = .26, Conf.(.082 $\leq \rho_I \leq$ .415) = .95, and the probability heuristics model, r = .272, Conf.(.097 $\leq \rho_I \leq$ .416) = .95. That is, people with a large span were more likely to draw conclusions that corresponded to the predictions made by these theories. This supports the idea that people with a large span are more likely to use a complex strategy such as mental models or the application of different heuristics.

Span was negatively correlated with matching, r = –.309, Conf.(–.466 $\leq \rho_I \leq$ –.133) = .95. For the atmosphere heuristic, the correlation was also negative, but it was not significant, r = –.102, Conf.(–.287 $\leq \rho_I \leq$ .103) = .95. This suggests that people with a smaller span were more likely to use a superficial strategy. However, the predictions of the superficial heuristics are similar to the initial models that would be constructed. So, people with a small capacity might simply be less thorough in their search for counterexamples, opting to create one model and to stick with the conclusion based on it. This idea is supported by a positive correlation, r = .38, between span and the number of "no valid conclusion" responses, consistent with the idea that people with smaller capacities are not constructing more than one model, which would be necessary to determine that there is no valid conclusion.

Discussion

The results of Experiment 1 suggest that people were using a complex strategy, such as mental models or probabilistic heuristics. Performance was better for syllogisms requiring fewer models. Also, people responded that there was no valid conclusion for some syllogisms, as much as 32% of the time for Figures 3 and 4. Simple heuristic strategies do not predict this.

The relation between span and reasoning was also examined. The operation span test was a significant predictor of reasoning. This is consistent with previous findings that syllogistic reasoning taxes working memory (Gilhooly et al., 1993, 2003).

In terms of mental models, there was support for the idea that the degree to which people used mental models varied. People with a larger span were more likely to use mental models than people with smaller spans. It is also possible that people with a small span could be using a simpler heuristic strategy, such as matching. However, the percentage of responses that corresponded to these strategies was less than the percentage that corresponded to the predictions made by mental model theory. If participants are divided into quartiles, based on span score, only 53% of the responses made by the lowest span group corresponded to the predictions made by matching, compared to approximately 80% that corresponded to the predictions made by mental model theory.

In terms of the probability heuristics model, it is important to keep in mind that the predictions for simple heuristic strategies generally correspond to the initial model constructed according to mental model theory as well as the predictions made by the probability heuristics model if people only applied the min heuristic. These results are consistent with the idea that people with smaller spans were less thorough in their use of probabilistic heuristics. That is, people with a smaller span were more likely to only apply one probabilistic heuristic, even when more were appropriate.

Overall, these results suggest that the people with a smaller span were using either mental models or probabilistic heuristics, but were not very thorough. It seems that often they constructed only a single model or only applied the min heuristic. This idea was examined further in Experiment 2.

EXPERIMENT 2

Experiment 1 showed a relation between working memory span and logical reasoning performance. The aim of Experiment 2 was to further investigate these findings. First, in Experiment 2 response times were collected. If someone is using a superficial heuristic, such as matching, then response times should be relatively rapid and not differ across problems. However, both mental model theory and the probability heuristics model predict that there should be a difference in response time depending on the number of models or probabilistic heuristics that are used. For example, it should take more time to construct and manipulate three models than one.

In addition, confidence ratings were gathered. If a person is using a superficial heuristic such as matching, then confidence ratings should not differ across syllogisms. A person should be similarly confident for every problem. Mental model theory predicts that confidence should vary across problems (Quayle & Ball, 2000). Quayle and Ball demonstrated that when evaluating syllogisms, people were more confident for valid than for invalid syllogisms. In addition, people should be more confident for one model than for multiple model syllogisms. That is, if a person realizes that there is more than one way to represent the premises, there should be less confidence in the conclusion.

Method

Participants

A total of 40 undergraduates from the University of Notre Dame and 40 undergraduates from Indiana University South Bend participated. Of these, 26 were male, and 54 were female. They participated in exchange for payment or partial course credit. All were native English speakers.

Materials and procedure

Operation span. This test was administered the same way as in Experiment 1 except for one difference. The set sizes ranged from 3 to 6 words; set size of 7 was eliminated because only 9 people correctly responded for the set size of 7 in Experiment 1.

Syllogistic reasoning. There were two differences from Experiment 1. First, response times were collected. Second, participants were asked to rate how confident they were by using a scale from 1 to 7, with a 1 indicating little confidence and a 7 indicating extreme confidence. For each syllogism, people were asked, "How confident are you (1 = not at all, 7 = extremely)?"

Interview. At the conclusion of the reasoning task, participants were asked questions, listed below, concerning the methods that they used to reach conclusions.

1. "Did you develop any rules or shortcuts as you progressed through the task?"
2. "Did you have a default answer (i.e., when in doubt, go with . . .)?"
3. "Describe how you reached conclusions."

Results and discussion

All p values are less than .05 unless reported otherwise. For correlation measures, a bootstrapping method (Efron & Tibshirani, 1998) with 1,000 bootstrap samples was used to correct for any influence of skewed distributions. Confidence intervals of 95% are reported for these analyses.

Accuracy

Accuracy scores are presented in Table 1. Overall, accuracy ranged from 20% to 88%. Performance was significantly different for syllogisms requiring different numbers of models, $F(2, 158) = 323.76$, $MSE = 0.021$. Performance for one-model syllogisms was significantly better than that for two-model syllogisms, $F(1, 79) = 348.52$, $MSE = 0.025$, which was better than that for three-model syllogisms, $F(1, 79) = 13.67$, $MSE = 0.013$.

Response times

Response times were different for syllogisms requiring different numbers of models, $F(2, 158) = 42.45$, $MSE = 33.280$ (see Table 1). Response times were faster for one-model syllogisms than for two-model syllogisms, $F(1, 79) = 14.47$, $MSE = 34.939$, which were faster than those for three-model syllogisms, $F(1, 79) = 39.43$, $MSE = 23.539$. Thus, people spent more time processing the problem when more models were required.

Confidence ratings

Confidence ratings[2] were different for syllogisms requiring different numbers of models, $F(2, 158) = 122.42$, $MSE = 0.164$ (see Table 1). Ratings were higher for one-model syllogisms than for two-model syllogisms, $F(1, 79) = 149.85$, $MSE = 0.219$, and three-model syllogisms, $F(1, 79) = 142.73$, $MSE = 0.190$. Ratings for two- and three-model syllogisms did not differ, $F(1, 79) = 3.19$, $MSE = 0.083$, $p = .23$. People were also more confident for valid ($M = 5.2$, $SD = 1.1$) than for invalid syllogisms ($M = 4.6$, $SD = 1.2$), $F(1, 79) = 92.44$, $MSE = 0.172$. This is consistent with the study conducted by Quayle and Ball (2000). Both the number of models

[2]Confidence ratings were significantly negatively correlated with response times, $r = -.21$.

required and the validity affected confidence. Thus, people were consciously aware of a difference in difficulty among the syllogisms.

Strategy use

Weighted scores. The weighted scores (see Table 2) showed the same pattern as that in Experiment 1. However, here, mental model theory and the probability heuristics model did not differ, $F < 1$. Both had larger weighted scores than atmosphere, $F(1, 79) = 94.16$, $MSE = 20.628$, and $F(1, 79) = 56.81$, $MSE = 32.416$, respectively. The weighted score for atmosphere was larger than the weighted score for matching, $F(1, 79) = 23.59$, $MSE = 4.313$.

Working memory capacity

The correlation between span and reasoning was significant, $r = .285$, Conf.$(.097 \leq \rho_I \leq .464) = .95$. The correlation between span and one-model syllogisms was small and not significant, $r = .145$, Conf.$(-.069 \leq \rho_I \leq .336) = 0.95$. The correlations between span and two- and three-model syllogisms were larger, as predicted. For two- and three-model syllogisms the correlations were significant, $r = .234$, Conf.$(.037 \leq \rho_I \leq .439) = .95$, and $r = .318$, Conf.$(.119 \leq \rho_I \leq .499) = .95$, respectively. Despite the fact that the correlation for one-model syllogisms was not significant, the pattern was consistent with that of Experiment 1. People with a larger span performed better on the reasoning task than did people with a smaller span. Also, a larger span seemed to be especially important when more models were required.

Working memory and strategy use

Correlations between span and responses predicted by the different strategies were conducted. The strongest correlations were between span and mental models, $r = .261$, Conf.$(.063 \leq \rho_I \leq .432) = .95$, and between span and the probability heuristics model, $r = .239$, Conf.$(.071 \leq \rho_I \leq .433) = .95$. The correlations for matching and atmosphere were not significant, $r = -.096$, Conf.$(-.308 \leq \rho_I \leq .130) = .95$, and $r = .016$, Conf.$(-.203 \leq \rho_I \leq .259) = .95$, respectively. These results are similar to those of Experiment 1, although the correlations for the simple heuristics were weaker here.

Working memory capacity and response times

Working memory span and response time were negatively correlated, $r = -.224$, Conf.$(-.405 \leq \rho_I \leq -.022) = .95$. The larger a person's span, the faster they responded. The relationship between span and response time was significant for one-model syllogisms, $r = -.316$, Conf.$(-.477 \leq \rho_I \leq -.130) = .95$. According to mental model theory, this result, along with a positive relation between span and accuracy, suggests that people with larger capacities are more efficient at constructing a single model.

The relationship between span and two-model syllogisms was significant, $r = -.261$, Conf.$(-.427 \leq \rho_I \leq -.090) = .95$, but it was not for three-model syllogisms, $r = -.104$, Conf.$(-.320 \leq \rho_I \leq .117) = .95$. As for multiple model syllogisms, the weaker correlations suggest that people with smaller spans were not trying to construct multiple models. If they were, then they should have spent much more time on these problems (based on how much slower they are, relative to people with larger spans, to construct a single model). These results are also consistent with the probability heuristics model, although instead of models, people would be dealing with the application of different numbers of probabilistic heuristics.

Working memory capacity and confidence

There were no significant correlations between working memory and confidence ratings.

Interview

The results from the interview give some insight into people's performance. The use of a *superficial rule or shortcut* was reported by 44% of the participants. Of these, 22 had span scores above the median, and 13 had scores below. This is consistent with previous work that showed that good reasoners were more likely to report an awareness of simple rules (Galotti, Baron, & Sabini, 1986).

One might think that this pattern is in contrast to the earlier results, which suggest that people with smaller working memory capacities were more likely to use superficial rules. However, further examination reveals that the rules reported by people were different from the theoretically outlined heuristics. For example, some participants reported that when both premises contained the word "some" or both contained "no" there was no valid conclusion.

A total of 54% responded that they used a *default answer*. Of these, over 70% used "no valid conclusion", while 16% used "some", and 7% used "no". This supports the idea that when people find a syllogism to be too complicated, they give up and respond with "no valid conclusion".

The two most popular methods used to reach conclusions were *grouping*, via mental imagery, and *verbal strategies*. Grouping the terms in the premises was used by 37% of participants. For example, based on the premise, "All lawyers are cyclists", one participant imagined lawyers holding briefcases while riding bicycles. Following this, they would add the information from the second premise to this image. Another 37% of the participants used verbal strategies. Some people said that they reworded the premises. Others tried to eliminate the middle (common) term in order to link the end terms. Substituting one term for another (usually for "all") was also used. In addition, 4% said that they read the premises over and over until they reached a conclusion, 9% reported the use of circles, and finally, 13% said that the answer seemed right, or that it just came to them.

These results are somewhat similar to those reported by Ford (1995). While that study used a think-aloud protocol rather than a post-task interview, there are some similarities. In that study, participants were classified as using either a verbal or a spatial strategy. In the current study, the participants who reported grouping (i.e., visuo-spatial processing) would be similar to Ford's spatial group, and the participants who reported using verbal strategies would be similar in both studies. Johnson-Laird and colleagues (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991) have argued that both verbal and spatial strategies are consistent with mental model theory.

GENERAL DISCUSSION

In two experiments, the relation between working memory span and syllogistic reasoning was examined. The data support the idea that a larger span is related to better reasoning. People with larger spans performed better and responded more quickly. Thus, working memory capacity seems to play a central role in syllogistic reasoning. In terms of strategy use, there was some support for both mental model theory and the probability heuristics model.

Assessment of theories

Mental models

The results of the current study are consistent with the idea that people use mental models during syllogistic reasoning. Predictions regarding accuracy of mental model theory were confirmed, such as the effects of figure and number of models. Moreover, response times showed that people spent more time reasoning when syllogisms required the use of more models. In addition, people were more confident in their responses to syllogisms that required the use of only one as opposed to multiple models.

If people are using mental models, a further question is whether people use them exclusively. Participants claimed to have developed their own shortcut rules as they progressed through the task (see also Galotti et al., 1986), even though some were ineffective or incorrect. An example of a rule was: "If both premises contain the word 'some' or 'no', then there is no valid conclusion." That is, it appears as though people created their own idiosyncratic rules based upon their experience with some of the syllogisms and were not relying on previously theorized heuristics. This is supported by the pattern of response times as the task progressed. Specifically, when the presentation order of the syllogisms was broken down into quartiles (e.g., the first 16 syllogisms, the next 16, etc.), people were faster to respond as the task progressed, $F(3, 237) = 68.96$, $MSE = 0.000$. There was a 16-s decrease in response times from the first ($M = 39$ s, $SD = 17$) to last quartile ($M = 23$ s, $SD = 10$). In addition, the use of these shortcuts was more prevalent among people with larger spans. This suggests that people, especially those with larger spans, may have begun the task using a complex strategy like mental models, but then might have developed a shortcut as the task progressed, leading to faster response times.

As for people with smaller spans, the results support the idea that they were also using mental models but this was more difficult for them. Compared to people with larger spans, people with smaller spans had a tendency to reach a conclusion based on the initial model predicted by mental model theory. There was a smaller correlation between span and performance for one-model syllogisms than for multiple model syllogisms. This means that people with smaller spans could solve one-model syllogisms as well as people with larger spans. However, if more than one model was required, people with larger spans did much better.

In terms of response times, people with smaller spans took longer to reach a conclusion, suggesting that they had more difficulty constructing mental models. Also, if the highest and lowest span quartiles are considered, the people with a larger span showed a larger difference in response time between one- and three-model syllogisms than did the people with a smaller span. This suggests that people with smaller spans could work with a single model, but, when multiple models were required, either they did not recognize that more than one model was needed, or it was too difficult for them to construct and coordinate multiple models.

In Experiment 2, there was no relationship between confidence ratings and working memory span. Thus, people with smaller spans were just as confident as people with larger spans. This result supports the idea that people with smaller spans are less thorough in their use of mental models. That is, if people with smaller spans typically only construct one model and do not recognize that more are possible, they should be just as confident in their response as someone who is using multiple models. In other words, people with a smaller span might inaccurately think that they are being thorough.

Some researchers have criticized mental model theory because they do not think that people search for counterexamples (Newstead, Handley, & Buck, 1999). Meanwhile others have conducted studies that support this idea (Bucciarelli & Johnson-Laird, 1999; Byrne & Johnson-Laird, 1990). The current study is consistent with an intermediate position, suggesting that people with larger spans are more likely to consider and construct multiple models than are people with smaller spans.

Probability heuristics model

In both Experiments 1 and 2, 76% of participants' responses were predicted by the probability heuristics model. This was fewer than what was predicted by mental model theory for both Experiments 1 and 2; however, when considering weighted scores there was basically no difference between mental model theory and the probability heuristics model. Although these experiments were conducted from a mental model perspective, the predictions of the probability heuristics model fared well, especially when weighted scores were considered.

The data from the interviews conducted in Experiment 2 were somewhat consistent with the probability heuristics model. Some people reported that they were aware of shortcuts, such as a simple rule. Some people took this further by indicating that they used simple rules to narrow down the choices. Following this, they would use a more complex strategy, such as mental models, to determine the final conclusion. This is similar to a verification strategy described by Chater and Oaksford (1999). This strategy includes testing the conclusion that is derived from the heuristics to make sure it is valid. They stated that this might be accomplished by using a more complex method, such as mental models. Unfortunately, people did not, or were not able to, describe these strategies during the interviews. Also, as mentioned earlier, many of the shortcuts involved rules created by participants that are unrelated to the rules proposed by psychological theories.

Overall, the results of the current study are generally consistent with the probability heuristics model. This model differs from mental model theory in that it emphasizes the use of heuristics and probabilistic judgements and gives representations such as mental models a lower priority, describing them as merely a way to verify a conclusion reached via heuristics.

Other strategies

Even though the use of mental models or probabilistic heuristics received the most support, the use of other strategies is possible. A total of 7 participants (9%) in Experiment 2 reported that they used circles of some sort (e.g., Euler circles). Of these, two stated that they tried, but found it to be too difficult. Ford (1995) described a group of participants that used Euler circles when they had access to pencil and paper. However, it is unclear how well people can do this when they do not have access to pencil and paper. Also, Euler circles and Venn diagrams could be classified as a type of mental model with particular properties and characteristics.

The current study also further refuted the idea that people exclusively use superficial heuristics. First, if people were only using superficial heuristics, then they would never respond that there was no valid conclusion. However, every participant selected "no valid conclusion" at least once. On average, people responded with "no valid conclusion" at least 25% of the time.

Second, the percentage of responses that corresponded to superficial heuristics was not very high, often around 50%. Thus, at best, the most that could be said is that people used a superficial heuristic half of the time, although even this estimate seems inflated. A superficial heuristic view is also hurt by the fact that response times and confidence ratings varied. If superficial heuristics were being used this should have remained relatively constant.

The only support for the idea that at least some people used superficial heuristics came from the data in Experiment 1. People were more likely to have answers that corresponded to superficial heuristics if they had a smaller span. However, this does not necessarily mean that people with smaller spans were using heuristics. The predictions of the superficial heuristics are very similar to the conclusions drawn from the initial model predicted by mental model theory. When the response time data are considered, it appears as though the low span people were using mental models. There was a negative correlation between span and response times. If people with smaller spans were using heuristics, they should have been responding faster than people with larger spans. Instead, they spent more time reasoning.

Conclusion

It appears that there is a clear relation between working memory span and syllogistic reasoning. People with larger spans perform better than those with smaller spans. In terms of strategy use, the predictions made by mental model theory were the most consistent with performance, although the probability heuristics model also fared quite well. Predictions made by mental model theory were confirmed, with participants performing better, faster, and more confidently for syllogisms requiring the use of one model than for multiple model syllogisms. Also, people with a smaller spans were less likely than people with a larger span to be thorough in their use of mental models. In addition, it appears as though people are able to develop new strategies or shortcut rules, which could be classified as a heuristic, as they gain experience with syllogisms.

REFERENCES

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge, UK: Cambridge University Press.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology*, *108*(2), 157–193.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*(3), 247–303.

Byrne, R. M. J., & Johnson-Laird, P. N. (1990). Remembering conclusions we have inferred. In J. P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 109–120). Amsterdam: Elsevier, North-Holland.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*, 191–258.

Conway, A. R. A., Tuholski, S. W., Shisler, R. J., & Engle, R. W. (1999). The effect of memory load on negative priming: An individual differences investigation. *Memory and Cognition*, *27*(6), 1042–1050.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, UK: Oxford University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422–433.

Efron, B., & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

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

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General*, *115*(1), 16–25.

Gilhooly, K. J. (1998). Working memory, strategies, and reasoning tasks. In R. H. Logie & K. J. Gilhooly (Eds.), *Working memory and thinking*. Hove, UK: Psychology Press Ltd.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory and Cognition*, *21*, 115–124.

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology*, *11*, 473–498.

Gilhooly, K. J., Logie, R. H., & Wynn, V. (2003). Syllogistic reasoning tasks and working memory: Evidence from a sequential presentation of premises. *Current Psychology*, *21*, 111–120.

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging*, *9*(3), 356–371.

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

Johnson-Laird, P. N. (1992). Syllog [Computer program]. Retrieved from http://www.cogsci.princeton.edu/~phil/syl.txt

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*, 1–62.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199–207.

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory and Cognition*, *27*(2), 344–354.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *53A*(4), 1202–1223.

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory and Cognition*, *29*, 1073–1080.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211–227.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.
Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.
Woodworth, R. J., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*, 451–460.

Original manuscript received 22 January 2002
Accepted revision received 11 September 2003
PrEview proof published online date/month/year