Memory & Cognition
2010, 38 (7), 883-892
doi:10.3758/MC.38.7.883

Working memory and insight in the nine-dot problem

Jason M. Chein, Robert W. Weisberg, Naomi L. Streeter, and Shaleigh Kwok
Temple University, Philadelphia, Pennsylvania

In the present article, we examine the contribution of working memory (WM) to solution of the nine-dot problem, a classic insight problem. Prior research has generally demonstrated a limited role for WM in the solution of insight problems, which are typically assumed to be solved without conscious planning. However, MacGregor, Ormerod, and Chronicle (2001) proposed an information-processing model that solves the nine-dot problem by relying on a visual WM mechanism, which they term *lookahead*. In the present research, we examine whether performance on the nine-dot problem is indeed predicted by WM capacity. The results indicated that spatial WM capacity predicted the tendency to draw lines outside the configuration of dots and predicted the solution on a hint-aided version of the problem. Furthermore, within those solving the problem, higher spatial WM capacity was also related to faster solutions. The results support the information-processing model and suggest a more essential role for WM and planning in insight problem solving than has previously been acknowledged.

Insight in problem solving has had a long history in psychology, and throughout most of that history, it has been a source of controversy (see Weisberg, 2006, chap. 6). Insight is contrasted with analysis as a mode of solving problems. Among the factors presumed to distinguish these two classes of problems is the degree to which each utilizes working memory (WM): Analytic problems putatively place heavier demands on WM than do insight problems. In a number of recent studies, the role of WM in the solution of a range of insight problems has been examined (Ash & Wiley, 2006; Fleck, 2008; Gilhooly & Murphy, 2005; Lavric, Forstmeier, & Rippon, 2000). Those studies lend some support to the distinction between insight and analysis as modes of problem solving but leave open the specific role that WM plays in the solution of insight problems. For example, both Gilhooly and Murphy (2005) and Ash and Wiley (2006) argued that executive components of WM may contribute to insight problem solving by supporting the allocation of attention but suggested little or no role for domain-specific capacities of WM. Meanwhile, both Lavric et al. (2000) and Fleck (2008) concluded that WM plays no role in insight problem solving.

In the present article, we examine the possible role of WM in another classic insight problem—the nine-dot problem—and show that individual differences in spatial WM capacity predict performance. Motivated by our findings, we later reconsider the extant literature and argue that the evidence points to a more central role for WM in some insight problems than has been previously acknowledged. Together, these findings lead us to question the notion that insight and analytic problems are distinguished by their relative dependence on working memory and, furthermore, to question the value of a sharp dichotomy between insight and analysis as modes of solving problems.

Analysis Versus Insight in Solution of Problems

An example of solving a problem through analysis is a knowledgeable person's solution to a long division problem: The individual knows a set of rules—an algorithm—that will produce the answer. In problems solved through analysis in which there is no familiar algorithm, such as the Towers of Hanoi, the individual may rely on heuristic search of the problem space in order to construct the solution. In each of these cases of solution through analysis, the individual makes more or less steady progress toward solution (Metcalfe & Wiebe, 1987).

In contrast, solving a problem through insight typically requires that one change the way in which the problem is conceptualized, through *restructuring* of the problem (Weisberg, 1995). Consider the marrying man problem: *A man in our town has married 20 women from the town. Bigamy is illegal in our town, and yet the man has broken no law. Explain*. Nearly everyone initially interprets the phrase *has married 20 women* as meaning that the man has been married to them. In order to solve the problem, one must realize that the phrase can mean something else: The man is the clergyman who presided over the women's marriage ceremonies. Therefore, the solution requires restructuring.

The finding that some problems are solved through restructuring provokes questions regarding the mechanisms through which restructuring comes about. In their original analyses of insight, the Gestalt psychologists attributed restructuring to spontaneous mechanisms, analogous to the changes in structure that occur when one examines

J. M. Chein, jchein@temple.edu

a reversible figure, such as the Necker cube (Weisberg, 2006, chap. 6). Similarly, Ohlsson (1992) proposed that restructuring in problem solving comes about as a spontaneous response to the person's reaching an impasse in his or her attempts to solve a problem. Such a state initiates a spread of activation in semantic memory. This process may result in the activation of concepts that will prompt the person to suddenly become aware of a new way of approaching the problem—an "Aha!" experience—that may bring with it a solution.

Insight and WM

Thus, in the traditional Gestalt view (e.g., Scheerer, 1963) and in more recent characterizations of solution through insight (Ohlsson, 1992), it is assumed that insight results from processes outside the person's conscious control. One prediction arising from this view is that planning should play a minimal role in the solution of insight problems. Any planning that might be initiated by the problem is, because of the problem's design, destined to result in failure and impasse. Only after the problem has been restructured, which is assumed to occur independently from conscious planning, will real progress be made. A corollary of that lack of planning is that working memory should not play a role in the solution of insight problems, since WM is critical in planning (Gilhooly, 2005; Shallice, 1982). In support of this prediction, Lavric et al. (2000) used a dual-task paradigm to examine WM in the solution of insight versus analytic problems. Lavric et al. had people solve insight problems (simplified versions of the candle and two-string problems; see Weisberg, 1995) or an analytic problem (a version of the Wason four-card task), while at the same time counting tones presented by a computer. Keeping count of the tones was assumed to require WM and was expected to interfere with performance on only the analytic (logic) problem. The results supported the predictions, which Lavric et al. took to mean that people solve insight problems without planning or use of WM.

There is, however, reason to expect that planning and WM might be involved in the solution of at least one insight problem: the nine-dot problem (Lung & Dominowski, 1985; Scheerer, 1963; Weisberg & Alba, 1981). The nine-dot problem (see Figure 1) is extremely difficult: Very few naive participants are able to solve it, even when given 100 attempts (Weisberg & Alba, 1981). Individuals attempting the problem rarely even draw lines outside the square, again testifying to its difficulty. The Gestalt analysis of this problem assumes that the individuals structure the situation as a square, and therefore, all lines are drawn within its boundaries. This makes a solution impossible (Scheerer, 1963). A solution requires that the situation be restructured, so that lines can be drawn outside of the box.

However, questions have been raised about the Gestalt interpretation of the nine-dot problem. First, Weisberg and Alba (1981) told participants that the only way to solve the problem was to draw lines outside the square shape. This instruction resulted in all of the participants drawing lines of that sort, but still, only a small minority solved the problem. Thus, although drawing lines outside the square is necessary to solve the problem, it is not sufficient (for similar results, see Kershaw & Ohlsson, 2004; Lung & Dominowski, 1985). The solution of this problem must accordingly involve processing that continues even after the individual knows to extend the lines beyond the box. Furthermore, evidence against the basic Gestalt interpretation comes from a series of experiments conducted by Kershaw and Ohlsson (2004), demonstrating that difficulty in the nine-dot problem arises from multiple factors, including perceptual, knowledge, and process-based limitations. Specifically, they found that solution rates increased modestly when training or instructions were given to alleviate one impediment to the solution of the nine-dot problem (e.g., the perceptual limitation assumed by Gestalt theory) but that a substantial increase in the solution rate was obtained only when the problem presentation combined aids to solution that addressed multiple sources of difficulty (e.g., perceptual and knowledge-based difficulties). Together, the above findings indicate that the solution of the nine-dot problem does not result from a single instance of restructuring.

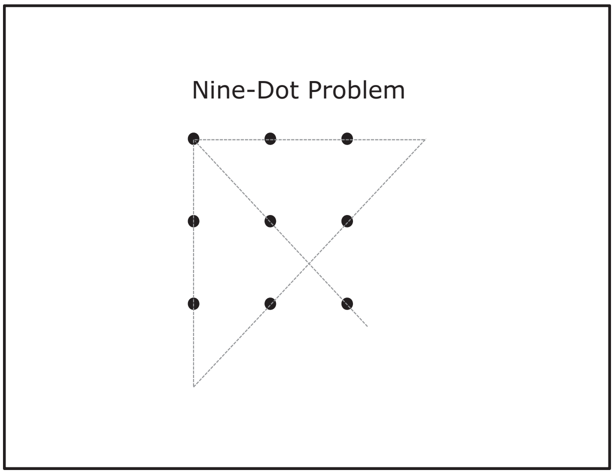

Figure 1. The nine-dot problem (a solution is shown with dotted lines). The instructions were to connect all of the dots by drawing four connected straight lines without lifting the pen from the paper. The problem was presented on a page with ample space in all margins, and the participants were given 10 min to complete the problem.

As an alternative to the Gestalt view, MacGregor, Ormerod, and Chronicle (2001; see also Chronicle, Ormerod, & MacGregor, 2001) proposed that the nine-dot problem is solved through the use of heuristic methods and that the difficulty in the problem is due to impediments in carrying out those heuristics based on limitations of WM. MacGregor et al. proposed that people working on the nine-dot problem try to draw lines that cover as many of the remaining dots as possible. A critical component in that strategy is mentally determining the situation that will remain after a series of lines has been drawn. That is, a critical component in people's ability to solve the nine-dot problem is *lookahead*—the person's ability to imagine in WM the result of carrying out various moves.

Chronicle et al. (2001) examined people's performance on specially designed dot problems, structured to either demand a high degree of lookahead or not; people solved the former problems more easily. Also, computer models designed with differing capacities for lookahead performed as predicted when given the nine-dot problem to solve (MacGregor et al., 2001). However, Chronicle et al. did not attempt to directly measure lookahead. The viewpoint of MacGregor et al. leads to the prediction that people with a greater ability to imagine and predict the outcomes of drawing lines—that is, people with greater lookahead skills—will differ in two ways from those with lesser ability: They will be more likely to draw lines outside the box, and they will solve the nine-dot problem more easily. In the present study, we tested those hypotheses, using spatial WM as a measure of lookahead.

EXPERIMENT 1

Method

Participants

Fifty-two Temple University undergraduate students participated in the study for course credit. Data from 1 student were incomplete and were therefore excluded from the analysis.

Procedure

The participants were tested individually in a single session lasting 1 h. All of the participants first attempted the nine-dot problem. Following the administration of the nine-dot problem, two computerized tests of WM were completed—one verbal and one spatial.

Nine-dot problem. The nine-dot problem was administered in its standard form. At the start of the session, the participants were seated at a table and given an unlimited number of worksheets, each containing a copy of the nine-dot problem. The participants were instructed that the solution of the problem required that all of the dots be connected using only four straight lines and that this goal had to be achieved without lifting the pen from the paper once an attempt was initiated. They were also instructed to "think on the paper," using additional sheets as needed, and to draw straight lines that passed through the center of each dot. Each participant was given 10 min to solve the problem.

Verbal WM task. Verbal WM capacity was measured using an automated version of the operation span (OSPAN) task (Unsworth, Heitz, Schrock, & Engle, 2005). In a typical OSPAN task, the participants must retain a series of verbal items (e.g., letters, words) that are presented in an interleaved fashion between a series of simple arithmetic equations. In the automated version of the task, the individual completes a number of practice arithmetic problems prior to WM assessment, and the mean time to solve those equations is used to titrate the rate of item presentation during WM testing. The to-be-remembered letters were subsampled from a set of 12 English consonants, and the participants completed three trials at each set size, ranging from 3 to 7 letters. Letter recall was tested at the end of each trial by displaying the complete array of 12 possible letters and requiring the participants to identify (by mouse click) the presented subset in serial order. The participants' final scores were calculated by summing the number of letters correctly identified (correct letter in the correct serial position) across all presented sets, with a maximum attainable score of 75.

Spatial WM task. Spatial WM capacity was measured using an automated version of the symmetry span (SSPAN) task adapted from Kane et al. (2004; see also Heitz & Engle, 2007), which followed the same structure as the automated measure of verbal WM. In the SSPAN task, the participants attempt to retain a sequence of spatial locations (positions on a 4 × 4 grid) presented in interleaved fashion between a series of symmetry judgments (in which the participants must determine whether the shaded regions of an 8 × 8 matrix are symmetric about the central vertical axis). The rate of item presentation is titrated to the individuals' speed at performing the symmetry judgment task alone. The participants again completed three trials at each set size, but the set sizes ranged from only two to five because of the relative difficulty of location memory. Location recall was tested at the end of each trial by displaying the 4 × 4 grid and requiring the participants to identify the presented subset of locations in serial order. The participants' final scores were the number of locations correctly identified (correct location in the correct serial position) across all presented sets, with a maximum attainable score of 42.

Results and Discussion

To assess the role of WM capacity in nine-dot performance, we identified high and low WM ability participants for each of the assessed WM measures (verbal and spatial) using median splits. As was noted above, the analysis of the nine-dot problem by MacGregor et al. (2001) predicts that lookahead capacity (operationalized here as spatial WM score) should be positively related not only to solution but also to the probability that the person will draw lines outside the grid formed by the dots. Of the 51 participants, only 7 drew lines outside the grid (14%). Despite those relatively few occurrences, the likelihood of drawing lines outside of the grid was significantly predicted by spatial WM performance, with the high spatial WM group accounting for 6 of the 7 occurrences [$\chi^2(1, N = 51) = 4.15, p = .04$]. Meanwhile, grouping the participants on the basis of verbal WM performance did not predict attempts to draw lines outside of the grid [$\chi^2(1, N = 51) = 0.65, p = .42$].

Treating WM as the dependent, rather than independent, variable yields similar support for a selective relationship between spatial WM and nine-dot performance. Specifically, the spatial WM scores were significantly greater for the group of 7 participants who drew lines outside of the grid ($M = 31.14, SD = 6.06$) than for those who did not ($M = 26.24, SD = 5.72$) [$t(49) = 1.99, p = .026$, one-tailed]. However, no such difference was demonstrated with verbal WM performance; the average OSPAN score for these 7 participants ($M = 57.29, SD = 12.05$) was actually slightly lower than the score for those whose attempts never extended outside of the box ($M = 62.12$, $SD = 9.23$), and the difference did not approach statistical significance [$t(49) = 1.23, p = .11$, one-tailed].

We next turned to the relationship between problem solution and WM performance. Only 5 of the 51 participants (~10%) went on to solve the nine-dot problem in the time allotted (i.e., 5 of the 7 individuals who drew lines outside the box subsequently solved the problem). Four of these 5 solvers were included in the high spatial WM ability group, whereas only 2 were included in the high verbal WM ability group. Although they are again suggestive of a possible role for spatial WM, but not verbal WM, in nine-dot solution, these patterns did not attain statistical significance by chi-square test [$\chi^2(1, N = 51) = 2.13, p = .14$, for high vs. low spatial WM ability; $\chi^2(1, N = 51) = 0.18, p = .67$, for high vs. low verbal WM ability]. Likewise, examination of WM performance for the solvers and nonsolvers indicated that the solvers performed better on the spatial WM task (Ms = 28.75 and 26.80, respectively), but the difference was not significant [$t(49) = 0.60, p > .05$]. No significant differences were found between the solvers and nonsolvers for verbal WM score (although the trend was again in the unexpected direction).

In conclusion, the results of this experiment were consistent in supporting a link between spatial WM and nine-dot performance: As was predicted by the analysis of MacGregor et al. (2001), we found a relationship between a measure of spatial lookahead and the probability that an individual would draw solution lines outside the box. The results concerning the relationship between WM and solution of the problem, although the trend was in the predicted direction, were not significant. However, our ability to directly demonstrate a relationship between lookahead and the subsequent solution of the nine-dot problem may have been limited by the small number of solvers. This outcome is not surprising, since the low proportion of solvers found in this experiment was comparable to that found in other studies of the nine-dot problem (e.g., Lung & Dominowski, 1985; Weisberg & Alba, 1981). Still, such low solution rates restricted statistical power and clearly limited our ability to adequately test the proposed relationship.

To address this limitation, in Experiments 2 and 3, we used nine-dot problem materials designed to yield higher solution rates than are typically observed with standard presentation of the problem. In those experiments, before the participants attempted the nine-dot problem, we administered training problems designed to demonstrate the necessity of extending lines past the implied grid in order to solve the problem. We also explicitly informed the participants that the solution would require extension of lines outside the box.

EXPERIMENT 2

In this experiment, participants were given hints designed to aid in the solution of the nine-dot problem, which allowed us to examine the relationship between WM performance and solution of the problem.

Method

Participants

Seventy-two Temple University undergraduate students received monetary compensation for participation. Fourteen reported having previously seen the nine-dot problem, and their data were excluded from analysis. The data were incomplete from 4 additional participants, who were also excluded from analysis.

Procedure

The participants were tested in pairs in a session lasting 1 h. The participants first completed the verbal and spatial WM measures and later attempted the nine-dot problem. The WM measures were administered as they were in Experiment 1. To increase the likelihood of nine-dot solution, we used a modified nine-dot procedure. Before attempting the nine-dot problem, the participants first completed four training problems (Figure 2) that together demonstrated the necessity of using an imaginary (off-grid) position as a vertex

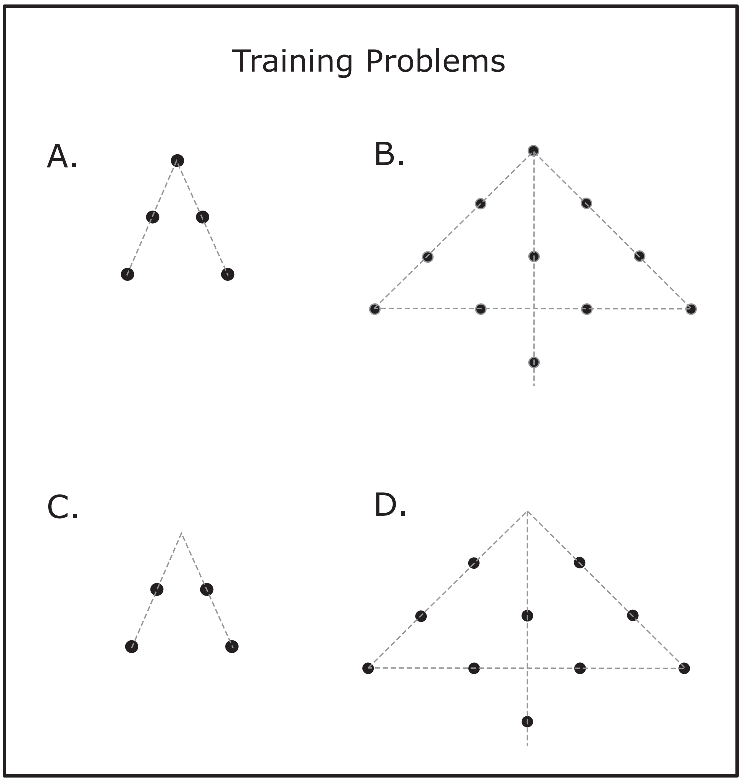

Figure 2. Training problems (the solution is shown with dotted lines) given prior to presentation of the nine-dot problem in Experiments 2 and 3. For Training Problems A and C, the participants were instructed to connect all of the dots by drawing two connected lines, and for Problems B and D, the participants were instructed to connect all of the dots by drawing four connected lines. The participants were given 3 min to solve Problems A and B and 5 min to solve Problems C and D.

Table 1
Working Memory (WM) and Nine-Dot Performance in Experiment 2: Correlations and Descriptive Statistics

	Variable							Descriptive Statistics			
Variable	1	2	3	4	5	6	7	*M*	*SD*	Range	Cronbach's α
1. Nine-dot solution	–							.46	.50	0–1	–
2. OSPAN recall (verbal WM)	.199	–						57.98	12.53	26–75	.86
3. SSPAN recall (spatial WM)	.452**	.328*	–					27.80	8.40	7–41	.80
4. Arithmetic accuracy	.072	.364*	.078	–				.92	.08	0–1	.87
5. Arithmetic reaction time	−.137	−.091	.004	−.222	–			1.17	3.66	0.7–2.2	.74
6. Symmetry accuracy	.193	.403*	.250	.116	.150	–		.88	.14	0–1	.82
7. Symmetry reaction time	−.081	−.354*	−.116	−.202	.262	−.063	–	6.48	1.24	0.4–10.2	.78

Note—$n = 54$. The α values were calculated as in Unsworth et al. (2005). $^{*}p < .05$. $^{**}p < .01$.

between successive lines in order to solve the problems. Kershaw and Ohlsson (2004) referred to this as a *nondot turn* and argued that this is the key element to success in the nine-dot problem (Kershaw & Ohlsson, 2001). After attempting the training problems, all of the participants were shown the correct solutions and the relationship among them: The solutions to Problems A and C are identical, as are the solutions to Problems B and D, but the latter problem of each pair includes an imaginary dot as a vertex.

To further facilitate a solution, all of the participants were also given the following hint prior to beginning the nine-dot problem: "The next problem is related to the problems that you just completed, and requires that you extend three lines past the dots in order to reach the solution." The participants had 10 min to solve the nine-dot problem.

For all of the problems (training and nine-dot), the participants had an unlimited number of worksheets, each containing a copy of the problem. The participants were again instructed to "think on the paper" using additional sheets as needed, to draw straight lines that passed through the center of each dot, and not to lift the pen from the paper once an attempt was initiated.

Results and Discussion

Descriptive statistics and pairwise correlations for all study variables are shown in Table 1. Nearly half of the participants (25 of 54, 46%) solved the modified nine-dot problem within the 10 min allotted, indicating that the training problems and hint were successful in promoting a substantially higher rate of solution than is observed in typical implementations of the nine-dot problem (see, e.g., Experiment 1, Weisberg & Alba, 1981). Importantly, it took most of the participants several minutes to reach the solution ($M = 317$ sec, $SD = 144$), suggesting that the aids to solution did not trivialize the problem.

To test the hypothesis that WM capacity predicts solution of the nine-dot problem, we again split the participants into high and low ability groups on the basis of independent median splits for performance on each of the WM tasks. The likelihood of a nine-dot solution was strongly linked to spatial WM ability. Of the 25 solvers, 19 were represented in the group of high spatial WM participants (only 6 in the group of low spatial WM participants), yielding a highly significant chi-square test [$\chi^2(1, N = 54) = 12.59, p < .001$]. In contrast, the likelihood of a solution was not related to verbal WM performance, with high and low verbal WM performance groups accounting for roughly equivalent proportions of the solvers (12 solvers in the high verbal WM group, 13 in the low verbal WM group) [$\chi^2(1, N = 54) = 0.74, p = .79$]. Table 2 includes a contrast of WM performance for the participants who solved the nine-dot problem with that for those who did not, which again indicates a difference in nine-dot performance based on spatial WM ability.

The larger number of solvers obtained in this experiment (than in Experiment 1) also permitted more powerful analyses of the correlations between WM performance and solution findings. Consistent with the results of the chi-square testing, a test of the simple correlation (point biserial) between spatial WM scores and nine-dot solution revealed a significant relationship [$r(52) = .452$, $p = .001$]. Meanwhile, the correlation between verbal WM (OSPAN) and nine-dot solution was considerably weaker and nonsignificant [$r(52) = .199, p = .148$], despite a significant correlation between spatial WM and verbal WM measures [$r(52) = .328, p = .016$]. The differential size of the simple correlations between spatial WM recall and nine-dot solution and between verbal WM recall and nine-dot solution was statistically significant on the basis of a Hotelling–Williams test of the difference between dependent correlations [$t(51) = 1.73, p = .046$, one-tailed].

To address possible confounding of the simple correlations because of shared variance between verbal and spatial WM task performance and because of the possible contributions of nonmnemonic task components, we conducted additional multiple regression analyses. In these analyses, we used logistic regression to accommodate the dichotomous coding of nine-dot solution. Our initial multiple regression model included spatial WM recall, verbal WM recall, and reaction time and accuracy from the processing components of each WM task (e.g., symmetry judgments, arithmetic operations) as simultaneous predictors of a nine-dot solution. Only spatial WM recall significantly predicted a nine-dot solution [odds ratio = 1.16, $p = .005$; for all other covariates, $p > .26$; omnibus model, $\chi^2(6, N = 54) = 15.4, p = .017$]. Since the bivariate correlations between verbal WM recall and pro-

Table 2
Working Memory (WM) Performance for Nine-Dot Solvers and Nonsolvers in Experiment 2

	Solvers		Nonsolvers		Solvers Versus Nonsolvers	
WM Measure	*M*	*SD*	*M*	*SD*	*t*(52)	*p*
OSPAN (Verbal WM)	60.88	7.96	56.72	15.17	1.5	.069
SSPAN (Spatial WM)	31.72	6.62	24.42	8.38	3.51	<.001

Note—$n = 54$; *t* tests were one-tailed.

cessing performance were stronger than those found between spatial WM and processing performance, we took the further precaution of repeating the logistic multiple regression with only spatial and verbal WM recall in the model (i.e., without the processing component predictors). This analysis produced results consistent with the larger model: Spatial WM recall significantly predicted a nine-dot solution (odds ratio = 1.16, $p = .004$), whereas verbal WM did not (odds ratio = 1.02, $p = .52$). The results from logistic multiple regression therefore corroborated the findings from simple correlation by again indicating a selective relationship between spatial WM and problem solution.

To further illustrate the selective role of spatial WM in supporting solution, we also examined the length of time taken to reach the solution among those who successfully solved the problem, with the expectation that individuals with higher WM capacity would solve the problem more quickly than those with relatively lower WM capacity. Using only the data from the subgroup of participants who solved the problem, we conducted median splits on the basis of each WM measure and performed between-groups tests of differences in the time taken to solve the problem (Table 3). A significant difference was found for only the spatial WM measure, with higher SSPAN solvers reaching a solution faster than lower SSPAN solvers. Verbal WM performance among the solvers was not significantly related to solution time.

Presenting the nine-dot problem with training problems and hints facilitated its solution (see Lung & Dominowski, 1985; Weisberg & Alba, 1981), and thereby permitted a stronger test of the hypothesis that solution of the nine-dot problem is related to WM capacity. Although it has been claimed that demand on WM distinguishes analytic from insight problem types, the present results strongly suggest that performance on at least the hint-aided version of the nine-dot problem is dependent on spatial WM capacity. This result might be presented as a challenge to earlier claims regarding the limited role of WM in insight problems. However, our use of revised materials to test the hypothesis that WM is related to solution of the nine-dot problem raises a concern: It might reasonably be argued that presenting the training problems and hints also alters the nine-dot problem so that it can no longer be considered an insight problem and, instead, more closely resembles problems solved through analysis. The purpose of Experiment 3 was to address this possibility.

Table 3
Time Taken to Solve the Nine-Dot Problem by Higher and Lower Working Memory (WM) Span Participants in Experiment 2

	High Span (sec)		Low Span (sec)		High Versus Low	
WM Measure	*M*	*SD*	*M*	*SD*	*t*(23)	*p*
OSSPAN (Verbal)	350	127	282	159	1.19	.120
SSPAN (Spatial)	264	155	367	119	1.86	.038

Note—$n = 25$; *t* tests were one-tailed.

EXPERIMENT 3

It has been shown that when asked to provide subjective judgments about their proximity to a solution (or *feelings of warmth*, FOW) throughout the solving process, individuals exhibit different patterns of FOW judgments when solving insight versus analytic problems (Metcalfe & Wiebe, 1987). Specifically, Metcalfe and Wiebe found that participants' FOW increased gradually for analytic problems, whereas for insight problems, FOW judgments increased relatively abruptly just before a solution was reached. Thus, it was proposed that the phenomenology of insight problems differed from that of noninsight problems, in accord with what one would expect on the basis of the notion that insight problems are solved in "Aha!" experiences. Furthermore, Metcalfe and Wiebe proposed that FOW ratings could be used to distinguish the two types of problems.

Accordingly, in Experiment 3 we collected warmth ratings from participants as they attempted the nine-dot problem after having been exposed to the training materials and hint and used those ratings to examine the phenomenology of this variant of the problem. If our hint-aided presentation of the nine-dot problem affected its status as an insight problem, we might expect to observe gradual increases in FOW over the protracted period of problem solving. On the other hand, if the revised presentation of the problem maintained the typical phenomenology of an insight problem, we would expect to find relatively sharp increases in warmth ratings in the period immediately preceding solution.

This experiment also has intrinsic interest beyond our probe into the role of WM in problem solution. To our knowledge, this is the first time that FOW ratings have been gathered from participants attempting to solve the nine-dot problem in any form. Therefore, this experiment for the first time provided information about the phenomenological characteristics of this classic problem.

Method

Participants

Twenty-seven Temple University undergraduate students participated in the study for course credit. Two reported having previously seen the nine-dot problem, and their data were excluded from analysis.

Procedure

The participants were tested individually in a single session lasting 1 h, which included additional measures that were not relevant to the present study. The nine-dot problem was administered as it was in Experiment 2 (with training problems and a hint), except that FOW ratings were gathered during the solution attempt on the nine-dot problem.

FOW ratings. We collected warmth ratings from the participants using a modified version of the procedure employed by Metcalfe and Wiebe (1987). Before beginning their nine-dot attempts, the participants were familiarized with a scale displaying the integers from −5 through 0 to +5. Every 15 sec, when a computer-generated tone sounded, the participants spoke aloud one of the numbers on the scale to indicate how close they felt they were to reaching the correct solution. The participants were told that in choosing 0 they would be indicating that they were neutral and had no sense or feeling, positive or negative, of how close they were to the solution. A choice of -5 meant that they were completely at a loss, and a choice of $+5$ meant they were almost certain that they knew the solution. The participants were also instructed to provide an FOW rating at the point of solution.

Results and Discussion

Consistent with the results from Experiment 2, just under half of the participants (12 out of 25; 48%) solved the nine-dot problem within the 10 min allotted ($M = 221$ sec, $SD = 136$).[1] FOW ratings for the last seven time points (accounting for the 90 sec leading up to the solution, and including the final rating given at the point of solution) were averaged for all of the solvers and were compared with the average final seven ratings provided by the 13 nonsolvers (Figure 3). All of the participants required at least 90 sec to reach a solution, thus motivating us to use this window of time for analysis.

As can be seen in Figure 3 (left panel), the FOW ratings provided by nonsolvers did not change during the last 90 sec in which they unsuccessfully attempted to solve the problem. In contrast, the participants who succeeded in solving the problem exhibited a pattern of ratings expected from an insight problem (Metcalfe & Wiebe, 1987): a sharp increase in FOW rating in the final moments leading up to the solution. Following a relatively lengthy period of flat FOW ratings (over 3.5 min on average), the solvers' ratings suddenly increased only 15 sec before the solution was completely drawn. The differences between the solvers' and nonsolvers' FOW ratings (solvers minus nonsolvers) are also shown in Figure 3 (right panel) and further illustrate the sharp increase in FOW ratings that corresponded with the penultimate rating (-15 sec). These FOW data were assessed statistically using a 2×7 ANOVA, with participant group (solvers vs. nonsolvers) as a between-subjects factor and time of rating (-90 to 0 sec) as a within-subjects factor.[2] The test yielded a predicted interaction between participant group and rating time [$F(1,20) = 20.60$, $p < .001$]. This interaction was probed further using planned contrasts comparing FOW ratings for the two participant groups at each time point and by conducting pairwise comparisons of FOW ratings given at each time point within the group of solvers. These contrasts showed a significant difference in average FOW ratings between solvers and nonsolvers for only the final two ratings [$t(20) = 2.96$, $p = .007$, for -15 sec; $t(15) = 7.27$, $p < .001$, for 0 sec]. Contrasts within the group of solvers further showed that only these final two ratings (-15 sec, 0 sec) were significantly higher than the earlier ratings [$t(9) > 2.54$, $p < .05$, for -15 sec; $t(4) > 2.58$, $p < .05$, for 0 sec], with the final rating (given at the solution) significantly higher than the penultimate (-15 sec) rating [$t(4) = 2.54$, $p = .022$].

In conclusion, the configuration of FOW ratings found here conforms to the pattern expected from an insight problem in showing that after minutes of working on the problem without any sense that progress had been made, the participants experienced a sudden sharp rise in FOW immediately preceding the point of solution. Those results are the first demonstration of the "Aha!" aspect of solving a variant of the nine-dot problem. In addition, these findings indicate that our modification of the format of the nine-dot problem resulted in an FOW rating profile that is consistent with the phenomenological status of an insight problem. With this validation of our methodology, in the

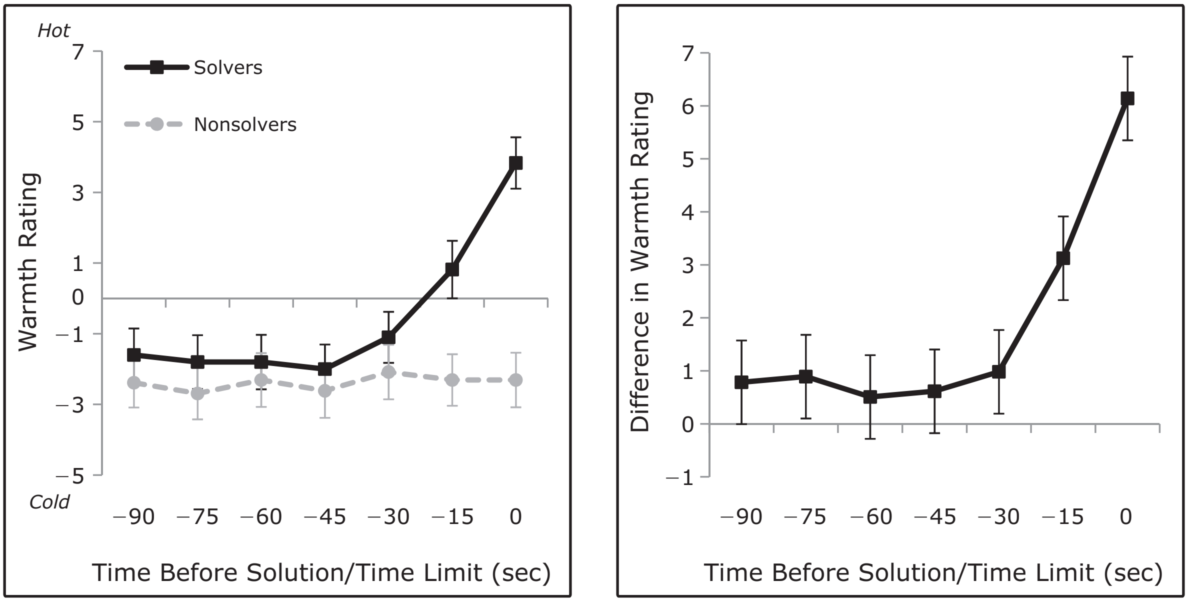

Figure 3. Feeling-of-warmth (FOW) ratings given by nine-dot solvers and nonsolvers in Experiment 3. Ratings were spoken aloud once every 15 sec during the problem attempt and were given on a scale ranging from -5 (*completely at a loss*) to $+5$ (*almost certain of the solution*). The left panel shows mean FOW ratings given for the final 90 sec spent working on the problem. For the solvers, the final rating was given upon reaching the solution, and for the nonsolvers, the final rating was given upon reaching the 10-min time limit. The right panel shows the mean difference in the solvers' and nonsolvers' ratings at each interval (solvers' FOW ratings minus nonsolvers' FOW ratings).

General Discussion section we reexamine prior findings and discuss the implications for claims about the role of WM in problem solving and, more generally, for theories that attempt to classify different problem types.

GENERAL DISCUSSION

The present experiments provided support for the analysis of the nine-dot problem of MacGregor et al. (2001) and, in so doing, raise important questions regarding the distinction between insight and analysis as modes of solving problems. Experiment 1 demonstrated that, as was predicted, people with higher spatial WM spans—that is, people higher in visual lookahead capacity—were more likely to draw lines outside the square formed by the dots (there was also a nonsignificant trend indicating that people with higher spatial WM capacity were more likely to solve the problem).

In order to conduct a stronger test of the predicted relation between lookahead and the solution of the nine-dot problem, we turned to the use of problem materials containing practice problems and hints, which have in the past produced an increase in solution probability (Lung & Dominowski, 1985; Weisberg & Alba, 1981). In Experiment 2, we found that participants with greater spatial WM capacity were more likely to solve the modified nine-dot problem. Moreover, for those who did solve the problem, individuals with higher spatial WM spans did so more quickly than those with lower spatial WM spans. Experiment 3 demonstrated that the use of hints to raise the base rate of solution of the nine-dot problem did not change the subjective phenomenology of the problem, in that the FOW ratings still exhibited patterns typical of insight problems. These results suggest that performance on at least one putative insight problem—assuming that the hint-aided nine-dot problem that we used is indeed an insight problem—may rely critically on the ability to plan out solution paths in WM.

In showing that problem solution was selectively related to spatial WM, and not to verbal WM, our findings shed further light on the particular WM mechanisms that may influence nine-dot problem solving. Namely, these results suggest that variation in performance is not explained by individual differences in executive (domain-general) aspects of WM (i.e., mechanisms that are common to verbal and spatial WM tasks). This finding additionally diminishes the likelihood that a third factor that correlates with executive WM, such as general fluid intelligence, might account for the relationship between WM and nine-dot performance. Instead, the results point to the importance of domain-specific mechanisms that support the ability to encode and maintain visuospatial representations in WM.

Although we interpret these findings as evidence that WM capacity specifically underlies nine-dot performance, we acknowledge that limitations of the available data prevent us from fully ruling out alternative explanations. For instance, we demonstrated a statistically significant relationship between WM capacity and problem solution only under the hint-aided conditions. It is thus possible that increased WM capacity simply allowed the solvers to use the hints during the problem attempt and was not specifically used to support more comprehensive lookahead, as we have suggested. However, this alternative interpretation does not readily explain the selective relationship observed between spatial WM capacity and problem solution. That is, it could be argued that verbal WM would be equally useful in maintaining a mental representation of the hint (i.e., in verbal form) during the problem attempt. Nor does this alternative explanation account for the differences in the character of problem performance (i.e., earlier drawing of lines outside the box by high spatial WM participants) that we observed in Experiment 1.

Another possible explanation of our findings does not implicate WM at all. Perhaps the selective relationship that we observed between nine-dot performance and the spatial WM task does not indicate the involvement of WM mechanisms per se but, rather, indicates a reliance on nonmnemonic spatial abilities. Of course, any such ability would have to be strongly correlated with the present measure of spatial WM capacity to yield the present results, thus rendering the distinction difficult to specify (and, indeed, most spatial abilities themselves rely on spatial WM; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Moreover, our data contain some evidence pointing to the specific involvement of spatial WM mechanisms: Although interparticipant differences in recall performance in the spatial WM task significantly predicted nine-dot performance, interparticipant differences in the speed and accuracy of symmetry judgments (the processing component of the spatial WM task) did not.[3] Although investigation of the relative involvement of mnemonic and nonmnemonic spatial abilities might help to further adjudicate between alternative interpretations, the available evidence leads us to the preferred interpretation that spatial WM capacity directly subserves nine-dot performance.

The present findings are in apparent contrast with those of prior studies suggesting that there is a limited relationship between domain-specific WM capacities and insight problem solving (Ash & Wiley, 2006; Fleck, 2008; Gilhooly & Murphy, 2005; Lavric et al., 2000). That apparent discrepancy can, of course, be reconciled by assuming that the nine-dot problem is (1) unique among insight problems in its reliance on spatial WM, (2) really an analytic problem, despite its traditional classification and a phenomenological relationship to insight problems, or (3) a hybrid problem with both analytic- and insight-problem characteristics (Weisberg, 1995). We note, however, that any of those positions would imply limited utility in dichotomizing problems into two general classes. Moreover, on the basis of a closer examination of prior studies (discussed below) we argue in favor of a fourth position: WM is important in the solution of many insight problems.

The present discovery of a link between WM and one putative insight problem (the nine-dot as used here) suggests that it may be prudent to reexamine the purported negative evidence concerning a relationship between insight and WM. Lavric et al. (2000) made perhaps the strongest assertion that there is no role for WM in insight problem solving. Their conclusion is potentially undermined, however, by a modality confound in their research between WM task and problem type: The analytic task was a verbal task, whereas the two insight problems were presented as visual problems. Since the concurrent tone-counting task was also a verbal task, disruption of only the analytic problem may have occurred because of competition for resources in the verbal domain and not because the problem classes rely differentially on the mechanisms of WM.

Although Gilhooly and Murphy (2005) were cautious in the interpretation of their large-scale individual-differences study, their results have also been taken as support for the notion that WM is involved in analytic, but not insight, problem solving (e.g., Fleck, 2008). This conclusion seems to stem primarily from the observation that performance on Raven's Progressive Matrices (Raven, Raven, & Court, 2003), a task known to correlate strongly with executive WM (Kane et al., 2004), accounts for a significant proportion of the variance in problems classified as analytic but not for those classified as involving insight. However, Gilhooly and Murphy reported simple correlations between Raven performance and problem solving that were statistically significant, and of similar magnitude ($r = .41$ for insight, $r = .50$ for analytic), for the two classes of problem. Perhaps more importantly, correlations with direct measures of WM capacity (sentence span, visual span) were also found to be significant, and equal, for the two classes of problem. Gilhooly and Murphy cautioned that simple correlations can reflect confounding with task-specific factors, but their WM measures explained a similar proportion of the variance for insight and noninsight problems even when multiple regression was conducted to take task-specific factors into account.

From the present work, we suggest that the strength of the relationship between WM and insight problem solving will depend on the specific nature of the WM and problem-solving tasks. Since Gilhooly and Murphy (2005) reported findings based only on composite indices of analytic and insight problem solving, possible problem-dependent differences in the engagement of WM in insight problem solving may have been obfuscated. Indeed, when data have been individually reported for other putative insight problems, such as matchstick problems (Ash & Wiley, 2006), the remote associates test (Ricks, Turley-Ames, & Wiley, 2007), or the "Lilies" problem (Fleck, 2008), the results often indicate significant correlations between WM and problem performance.

In conclusion, despite previous claims regarding the limited role of WM in insight problem solving, the results from the present study and from several other recent studies appear to provide consistent evidence that WM ability predicts success on insight problems. If so, one interesting question pertains to at which phase of the problem WM makes its critical contribution: Is it during initial problem search, restructuring, or postrestructuring operations? To partly address this question, Ash and Wiley (2006) developed insight problems having either an extensive or a limited initial search space (thus isolating the restructuring component) and examined correlations with WM. WM capacity most strongly predicted success on problems with a large initial search space, leading Ash and Wiley to conclude that WM is essential to the initial problem search. In our modified implementation of the nine-dot problem (Experiment 2), participants were instructed that the solution could not be achieved without extending the lines beyond the square area defined by the nine dots. Those instructions may have allowed the participants to bypass the initial search and restructuring steps that are a part of typical nine-dot presentation. Accordingly, our findings seem to complement those of Ash and Wiley, by indicating that WM must play an important role not only before initial restructuring has occurred, but also afterward.

In summary, our results have both narrow and broad implications. First, we provide support for MacGregor et al.'s (2001; Chronicle et al., 2001) analysis of the nine-dot problem and show that visual lookahead, as operationalized through a spatial WM measure, plays a critical role in its solution. Second, our results (and our reinterpretation of the results of several recent studies) raise a broader question concerning the distinction between analysis and insight as modes of problem solving, especially when that distinction is based on the putative role of WM or the lack thereof in problem solution. It has been argued elsewhere (Fleck & Weisberg, 2004; Perkins, 1981; Weisberg, 1995, 2006, chap. 6) that a fine-grained analysis of the solution of insight versus analytic problems indicates that the processes underlying solution of those two putative types of problems overlap in many ways; we believe that the present results support this conclusion. Therefore, although the notion of insight as a distinct process has a long history in the psychological study of problem solving, it might be useful at this point to refrain from using *analytic* and *insight* as theoretical terms applied a priori to problems. Rather, those terms should only be applied as a result of analyses of data indicating clear distinctions between problem types. The present analysis indicates, however, that such differences may be difficult to find.

AUTHOR NOTE

The authors thank Jennifer Andrade and Ryan Oliver for their efforts in support of data collection and analysis. Correspondence concerning this article should be addressed to J. M. Chein, 825 Weiss Hall, 1701 N. 13th Street, Department of Psychology, Temple University, Philadelphia, PA 19122 (e-mail: jchein@temple.edu).

REFERENCES

ASH, I. K., & WILEY, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review*, **13**, 66-73.

CHRONICLE, E. P., ORMEROD, T. C., & MACGREGOR, J. N. (2001). When insight just won't come: The failure of visual cues in the nine-dot problem. *Quarterly Journal of Experimental Psychology*, **54A**, 903-919.

FLECK, J. I. (2008). Working memory demands in insight versus analytic problem solving. *European Journal of Cognitive Psychology*, **20**, 139-176.

FLECK, J. I., & WEISBERG, R. W. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory & Cognition*, **32**, 990-1006.

GILHOOLY, K. J. (2005). Working memory and planning. In R. Morris & G. Ward (Eds.), *The cognitive psychology of planning* (pp. 71-88). London: Routledge.

GILHOOLY, K. J., & MURPHY, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, **11**, 279-302.

HEITZ, R. P., & ENGLE, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, **136**, 217-240.

KANE, M. J., HAMBRICK, D. Z., TUHOLSKI, S. W., WILHELM, O., PAYNE, T. W., & ENGLE, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, **133**, 189-217.

KERSHAW, T. C., & OHLSSON, S. (2001). Training for insight: The case of the nine-dot problem. In J. D. Moore & K. Stennigs (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 489-493). Mahwah, NJ: Erlbaum.

KERSHAW, T. C., & OHLSSON, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 3-13.

LAVRIC, A., FORSTMEIER, S., & RIPPON, G. (2000). Differences in working memory involvement in analytical and creative tasks: an ERP study. *NeuroReport*, **11**, 1613-1618.

LUNG, C.-T., & DOMINOWSKI, R. L. (1985). Effects of strategy instructions and practice on 9-Dot problem-solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 804-811.

MACGREGOR, J. N., ORMEROD, T. C., & CHRONICLE, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 176-201.

METCALFE, J., & WIEBE, D. (1987). Intuition in insight and noninsight problem-solving. *Memory & Cognition*, **15**, 238-246.

MIYAKE, A., FRIEDMAN, N. P., RETTINGER, D. A., SHAH, P., & HEGARTY, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, **130**, 621-640.

OHLSSON, S. (1992). Information processing explanations of insight and related phenomena. In M. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1, pp. 1-44). London: Harvester-Wheatsheaf.

PERKINS, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.

RAVEN, J., RAVEN, J. C., & COURT, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary scales: Section 1. General overview*. San Antonio: Harcourt Assessment.

RICKS, T. R., TURLEY-AMES, K. J., & WILEY, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition*, **35**, 1456-1462.

SCHEERER, M. (1963). Problem solving. *Scientific American*, **208**, 118-128.

SHALLICE, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B*, **298**, 199-209.

UNSWORTH, N., HEITZ, R. P., SCHROCK, J. C., & ENGLE, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, **37**, 498-505.

WEISBERG, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 157-196). Cambridge, MA: MIT Press.

WEISBERG, R. W. (2006). *Creativity: Understanding innovation in problem solving, science, invention, and the arts*. Hoboken, NJ: Wiley.

WEISBERG, R. W., & ALBA, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General*, **110**, 169-192.

NOTES

1. The data from 1 solver were excluded from further analysis because he did not realize that he had reached the solution, and a follow-up interview indicated that his criteria for solution differed from those indicated in the instructions.

2. Because of an error in procedure, the final FOW ratings from the time of solution were not recorded from 5 participants.

3. Since symmetry and operation judgments were made concurrent with WM recall demands, it is possible that the participants compensated for difficulty with these processing tasks by sacrificing recall accuracy. Such a strategy could have weakened possible correlations with nine-dot performance. The automated WM tasks also assess single-task processing performance during the initial practice trials. We therefore repeated logistic multiple regression analyses using WM recall and single-task processing performance (accuracy and reaction time) as simultaneous predictors. The results again indicated significant correlations between only spatial WM recall and nine-dot problem solution.

(Manuscript received August 11, 2009;
revision accepted for publication April 12, 2010.)