Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 1, 3–13


0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.1.3

# Multiple Causes of Difficulty in Insight: The Case of the Nine-Dot Problem

Trina C. Kershaw and Stellan Ohlsson
University of Illinois at Chicago

Theories of insight problems are often tested by formulating hypotheses about the particular difficulties of individual insight problems. Such evaluations often implicitly assume that there is a single difficulty. We argue that the quantitatively small effects of many studies arise because the difficulty of many insight problems is determined by multiple factors, so the removal of 1 factor has limited effect on the solution rate. Difficulties can reside either in problem perception, in prior knowledge, or in the processing of the problem information. We support this multiple factors perspective through 3 experiments on the 9-dot problem (N.R.F. Maier, 1930). Our results lead to a significant reformulation of the classical hypothesis as to why this problem is difficult. The results have general implications for our understanding of insight problem solving and for the interpretation of data from studies that aim to evaluate hypotheses about the sources of difficulty of particular insight problems.



After a period of relative neglect, the problem of insight is again attracting the attention of researchers (Ansburg & Dominowski, 2000; Chronicle, Ormerod, & MacGregor, 2001; Kaplan & Simon, 1990; Kershaw & Ohlsson, 2001; Kershaw, Ohlsson, & Coyne, 2003; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Knoblich, Ohlsson, & Raney, 2001; MacGregor, Ormerod, & Chronicle, 2001; Metcalfe, 1986; Metcalfe & Wiebe, 1987; Ohlsson, 1992; Ormerod, MacGregor, & Chronicle, 2002; Schooler & Melcher, 1995; Schooler, Ohlsson, & Brooks, 1993; Smith, Ward, & Finke, 1995; Sternberg & Davidson, 1995). Although some researchers have argued that insight problem solving is no different from analytical problem solving (Weisberg, 1986; Weisberg & Alba, 1981), the distinction is strongly supported by meta-cognitive results (Metcalfe & Wiebe, 1987), the effects of verbal overshadowing (Schooler et al., 1993), and neuroscience findings (Beeman, Haberman, & Bowden, 2002; Bowden & Beeman, 1998; Lavric, Forstmeier, & Rippon, 2000).

Insight problems are objectively simple in that the problem statements are short, the problem situations contain only a handful of object and relations, and the solutions require few actions. Yet, they are difficult to solve. One approach to understanding insight problem solving is to pinpoint the sources of difficulty. Hypotheses about difficulty can be tested by altering the problem (MacGregor et al., 2001), giving hints (Ormerod et al., 2002), and providing training that either facilitates (Gick & Holyoak, 1980) or hinders (Birch & Rabinowitz, 1951) problem solving. The logic behind these three methods is the same: If the difficulty of problem X is due to feature Y, then removing that feature or lowering its importance should raise the solution rate to a level close to 100%.

However, this strong prediction has seldom been supported in empirical studies. Table 1 summarizes a century's worth of attempts to alleviate the difficulty of the nine-dot problem (connect nine dots arranged in a $3 \times 3$ square with exactly four straight lines without lifting the pen from the paper and without retracing; Maier, 1930; see Figure 1). The expected solution rate for this problem under laboratory conditions (e.g., a time limit of a few minutes) is 0% (MacGregor et al., 2001). As Table 1 shows, there are only three classes of manipulations that have succeeded in raising the solution rate above 50%. The first is to simplify the problem by giving away the first or second line (Weisberg & Alba, 1981) and the second is to add more dots (MacGregor et al., 2001). Both of these alter the problem significantly. Lung and Dominowski (1985) provided strategy training and practice that raised the solution rate to 59%. Even this manipulation left a remainder of 41% failures. The situation is similar with respect to other classical insight problems.

The hypotheses tested in these and other studies attempt to pinpoint a single source of difficulty. For example, Duncker's candle problem is widely believed to be difficult because people tend to overlook that the thumb tack box can be converted into a candle holder, and the nine-dot problem is believed to be difficult because people are seduced into assuming that all lines must be drawn inside the square formed by the dots (see Kershaw et al., 2003, for additional examples of single-difficulty hypotheses). The single-difficulty assumption implies that the weak effects of problem variants, hints, or training procedures are due to erroneous identification of the relevant difficulty or ineffective administration of the relevant manipulation.

The purpose of this article is to put forward an alternative interpretation: The difficulties of insight problems tend to be resistant to facilitation maneuvers because they are *overdeter-*

Trina C. Kershaw and Stellan Ohlsson, Department of Psychology, University of Illinois at Chicago.

The experiments contained within this article were completed as part of the requirements for the master's degree for Trina C. Kershaw. We thank Colleen Coyne, Martin Bis, Cherie Siu, and Patty Johnson for their assistance in collecting data and analyzing protocols. We also thank Jennifer Wiley and Andrew Conway for their helpful comments and Ivan K. Ash for his many insight conversations. In addition, we thank Tom Ormerod and an anonymous reviewer for their comments on a draft version of this article.

Correspondence concerning this article should be addressed to Trina C. Kershaw, University of Illinois at Chicago, Department of Psychology (M/C 285), Chicago, IL 60607. E-mail: tkersh1@uic.edu

Table 1
*Summarized Difficulties and Manipulations of the Nine-Dot Problem*

| Theorist | Source of difficulty | Manipulation | Solution rate (%) |
|---|---|---|---|
| Burnham & Davis (1969) | sequence of lines is not obvious | 1. outside permitted | 27 |
| | | 2. outside & dot to begin | 20 |
| | | 3. cross lines | 7 |
| | | 4. cross lines & dot to begin | 13 |
| | | 5. 11-dot | 43 |
| Maier & Casselman (1970) | need to extend lines beyond dots | nine dots placed within a box | 40 |
| Weisberg & Alba (1981) | past experience with dot problems unhelpful | 1. must go outside | 20 |
| | | 2a. must go outside & 1st line given | 62 |
| | | 2b. must go outside & 1st, 2nd line | 100 |
| | | 3. training | |
| | | a. connect dots | 0 |
| | | b. go "outside" | 43 |
| Lung & Dominowski (1985) | lines must begin/end on dots | 1. strategy | 34 |
| | | 2. practice | 22 |
| | | 3. strategy + practice | 59 |
| MacGregor et al. (2001) | problem abstractly defined; amount of lookahead determines solution rate | 1. 9-dot | 0 |
| | | 2. 11-dot | 50 |
| | | 3. 12-dot | 60 |
| | | 4. 13-dot | 70 |
| Chronicle et al. (2001) | Gestalt/perceptual properties are important | 1. shading | 16 |
| | | 2. unfilled circles | 33 |
| | | 3. instructions + 1 | 30 |

*mined*. That is, they have multiple causes. The strong intuition that there is a single difficulty is due to a conceptual confusion between the behavior that is required for a successful solution and the cognitive factors that determine the probability that this behavior comes to mind. In many insight problems there is indeed a single action or behavior that can be singled out as the key to the solution. Such key behaviors include making a pair of pliers into a pendulum in Maier's two-string problem and arranging the matchsticks in three-dimensional configurations in the six-matches problem. However, we propose that even where we can identify a single key behavior, the difficulty of a problem is determined by multiple factors that make the emergence of that behavior unlikely in the given problem situation.

We distinguish three classes of difficulty factors. The Gestalt psychologists emphasized *perceptual factors* (cf. Köhler, 1971/1930). Figural integrity, figure–ground relationships, and other Gestalt laws can influence problem perception in such a way that the crucial affordances inherent in the problem situation are overlooked. In contrast, researchers in the information-processing tradition (Newell & Simon, 1972; Kotovsky & Simon, 1990) focused on *process factors.* These include the size of the search space for the problem, the branching factor, the length of the solution, the specificity of the goal state (Burnham & Davis, 1969; Greeno, 1978; Simon, 1973), as well as the amount of mental lookahead needed to find the solution (MacGregor et al., 2001). In our work, we have emphasized *knowledge factors* (Kershaw & Ohlsson, 2001; Knoblich et al., 1999; Knoblich et al., 2001; Ohlsson, 1984b, 1992; see also Weisberg, 1986). Prior knowledge and experiences that are activated by the problem situation but that are not in fact helpful might place inappropriate constraints on the solution, thus lowering the probability that the key action comes to mind. The key point for our present purposes is that these different types of factors are mutually compatible. All three types of factors might be operating with respect to one and the same problem, or, more importantly, with respect to one and the same action.

The strong logic of single-difficulty explanations does not apply to problems with multiple sources of difficulty. The elimination of one factor is not enough to allow an individual to solve a problem. Manipulations that counteract individual sources of difficulty will raise the solution rate of an insight problem a significant but small amount. Alleviating one source of difficulty helps, but only by so much. To raise the solution rate higher, we must alleviate multiple sources of difficulty.

To apply this perspective to the nine-dot problem, we first need to identify the key behavior in the solution with some precision. The Gestalt psychologists (cf. Maier, 1930; Maier & Casselman, 1970; Scheerer, 1963) emphasized the good figure quality of the square formed by the dots and conceptualized the key behavior as *drawing lines outside the figure*. Lung and Dominowski (1985) suggested the closely related but nevertheless distinct formulation that lines must begin and end on dots. An examination of MacGregor et al.'s (2001) results suggests an additional reformulation. Two of the nine-dot variants designed by MacGregor et al. differed by only one dot. However, the solution rate for the 13-dot variant (see Figure 2) was 88%, whereas the solution rate for the 12-dot variant (see Figure 2) was 27%. In the 12-dot version, the dot was missing at the upper left corner of the problem, thereby requiring participants to turn on the white space on the paper instead of on a dot. We refer to turns in a place where there is no dot as a *non-dot turn*, and we suggest that this is the key behavior for success on the nine-dot problem (Kershaw & Ohlsson, 2001). The non-dot turns hypothesis incorporates the insights of the Gestalt psychologists and of Lung and Dominowski (1985), but the slight change in conceptualization becomes important when we consider the factors that might contribute to lowering the probability of this behavior.

All three types of difficulty factors might be operating in the nine-dot problem. First, making non-dot turns requires that one breaks with the good figure of the square, as the Gestalt psychologists pointed out, but one must also break with the figure–ground organization of the problem and see the white background of the

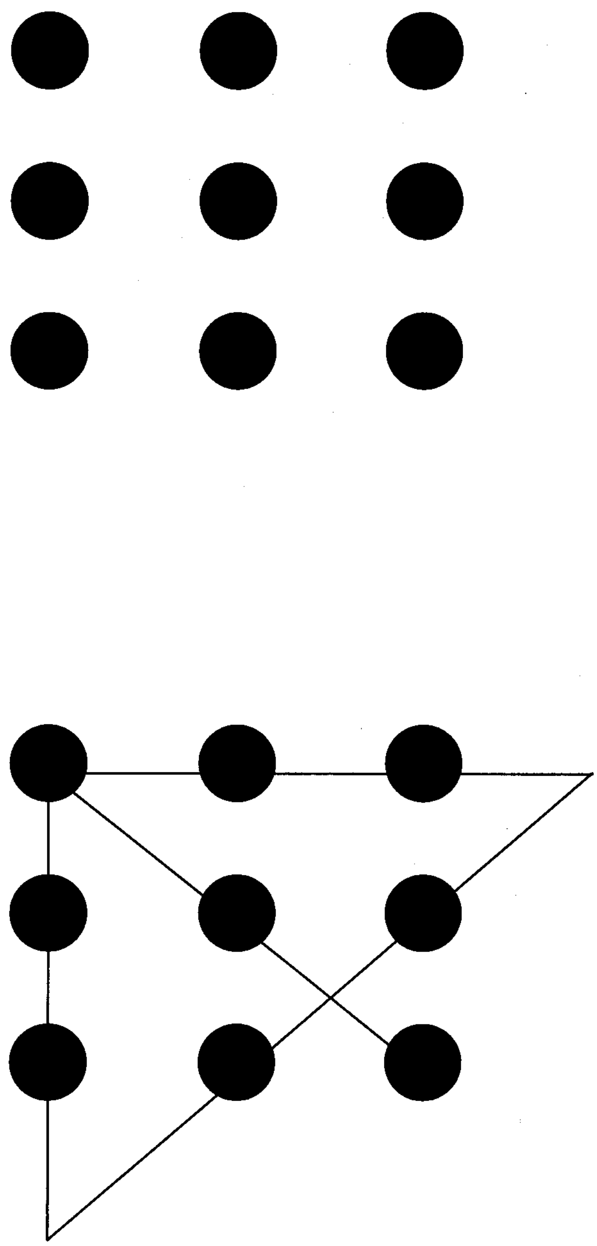

*Figure 1.* The nine-dot problem and its solution (Maier, 1930).

paper as available work space. Second, making a non-dot turn is more hindered than facilitated by the prior experience of the typical experimental participant. As previously noted by Weisberg and Alba (1981), the experience that people have with connecting dots might come primarily from children's connect-the-dots puzzles, in which lines are always drawn directly from dot to dot. Drawing lines on graph paper is another prior experience that might inappropriately constrain the set of lines considered. Third, as Burnham and Davis (1969) pointed out, it is not obvious where to start drawing the first line, the space of possible or legitimate lines is ill-defined once one goes outside the square, it is not obvious in which order the four lines should be drawn, and the goal state (the visual appearance of the correct configuration of lines) is only known via its abstract specification (four straight lines). MacGregor et al. (2001) suggest that to search the space of possible line configurations to a depth of four—the amount of lookahead needed to find the solution—is difficult due to capacity limitations on working memory. Consistent with this, the strongest facilitation maneuver to date was to provide participants with the first and second lines (Weisberg & Alba, 1981), which dramatically shrinks the search space. We conclude that perceptual, knowledge, and process factors might all be operating to suppress the probability of non-dot turns in the nine-dot problem.

In previous work (Knoblich et al., 1999), we argued that once a problem solver has overcome a knowledge factor—for example, relaxed an unhelpful constraint-like turn where there is a dot—there is no reason why that factor should continue to operate. When a knowledge factor is the sole source of difficulty, we expect the difficulty to disappear once the key action has been performed.

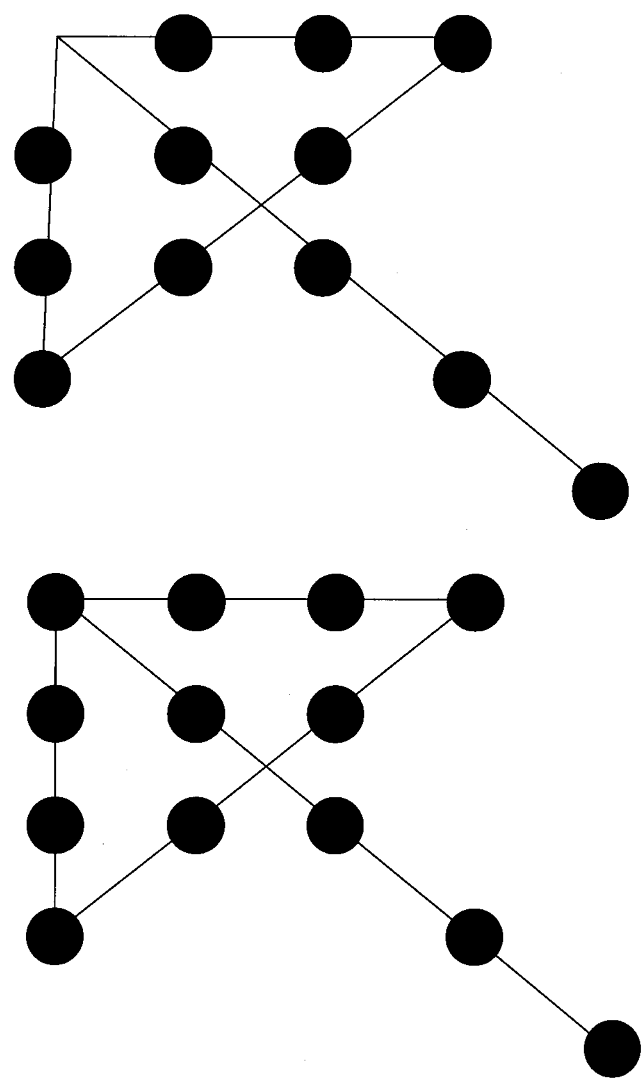

*Figure 2.* The 12-dot and 13-dot problems with their solutions (MacGregor et al., 2001).

Hence, if there is a need to perform the key action a second or third time, this need not increase problem difficulty. In the matchstick arithmetic domain, we found this to be true (Knoblich et al., 1999). When there are perceptual and process factors operating as well, this prediction does not follow. Organizational principles like good Gestalt and figure–ground are entrenched dispositions that will not cease to operate within the short time span of an experiment. Furthermore, even if the key action comes to mind, process factors might prevent it from being performed correctly. It is not enough to make non-dot turns, they have to be in the right place. Hence, when perceptual and process factors are operating, a need to perform the key action multiple times can be expected to increase problem difficulty.[1]

This analysis generates four predictions:

1. The more non-dot turns a dot problem requires, the more difficult the problem will be to solve.

2. Different single-difficulty manipulations of radically different natures, which address perceptual, knowledge, or process factors, can yield statistically significant effects on the solution rate.

3. Even when single-difficulty manipulations have significant effects, the size of the effects is small.

4. Multicomponent manipulations that address multiple difficulty factors will raise the solution rate more than single-difficulty manipulations.

Experiments 1 and 2 test Predictions 2 and 3, while Experiment 3 tests Predictions 1 and 4.

## Experiment 1

The purpose of Experiment 1 was to test two potential difficulty factors in the nine-dot problem: prior knowledge of non-dot turns and the figure–ground distinction. If the lack of prior experience of drawing non-dot turns is one of the factors that affect the difficulty of the nine-dot problem, then providing such experience should influence the solution rate. Experiment 1 compared facilitating (turning on a non-dot point) to hindering (always turning on a dot) training.

To draw non-dot turns, the problem solver has to bridge the gap between figure (the pattern of dots) and ground (the white part of the paper) and has to draw lines that turn on the ground. Helping problem solvers to bridge that gap should therefore influence the solution rate. However, Chronicle et al. (2001) found no effect of providing circles (unfilled dots) in the places where the non-dot turns needed to be made. We attempted to strengthen this manipulation by presenting the target problem embedded in a grid of circles (unfilled dots) that takes up an entire page, thus transforming the background into figure.

### *Method*

#### *Participants and Design*

Participants were 162 undergraduates from the University of Illinois at Chicago's (UIC) participant pool. Two participants were dropped from the final analysis because of prior knowledge of the nine-dot problem. No demographic data were collected about the participants.

The design of Experiment 1 was a 2 × 2 factorial. The two independent variables were type of training (facilitating vs. hindering) and the perceptual manipulation (grid vs. no grid). There were 40 participants in each of the four training groups. All participants attempted the same target problem, the nine-dot problem (see Figure 1).

#### *Materials and Procedure*

The participants were seen in groups of 2–10 students. They were given general instructions and a 24-page booklet, which contained instructions, the training exercises, the target problem, and screening for prior knowledge of the nine-dot problem. They then worked through the booklet at the pace indicated by the experimenter.

The participants in the facilitating groups solved 13 training problems that required them to make one, two, or three non-dot turns in the context of groups of black dots approximately 1 cm in diameter (see Figure 3B). The order of the training problems was the same for all participants. The participants had 1 min to complete each of the 12 first training problems.

The 13th problem (from MacGregor et al., 2001; Experiment 3, Problem 12b rotated 90 degrees) contained 12 dots, required one non-dot turn, and served as an opportunity to transfer the concept of a non-dot turn to a more complex dot configuration. Unlike the preceding practice problems, the 13th problem was presented on two successive pages. The participants were told to try out their solutions on the first page. When they had found a solution, they were to turn the page and redraw their solution on the second page. The participants had 4 min to complete the 13th practice problem (not including the time to redraw their solution). The participants then worked on the nine-dot problem, using the same 4-min, two-page procedure as for the 13th practice problem. At the end, the participants were asked to indicate on the last sheet whether they had seen the nine-dot problem before (prior to the experiment) and, if so, whether they remembered the solution.

For the participants in the hindering groups, the 13th training problem was presented with a dot inserted where the turn was needed (see Figure 3C). The procedure was in other respects the same as for the facilitating groups.

For the participants in the plain background groups, the black dots appeared on a plain background, as in Figure 3B and 3C. For the participants in the grid background groups, the dot groups of the first 12 training problems were presented superimposed on a grid of open circles of the same size as the black problem dots (see Figure 3A). They were instructed that they could use the open circles in their solutions, but that the goal for each problem was to connect the black dots. The 13th problem was presented on a plain page for all groups.

### *Results*

The solution rates for the four experimental conditions were as follows: facilitating grid (FG), 17.5% (7/40); facilitating no grid (FNG), 12.5% (5/40); hindering grid (HG), 2.5% (1/40); hindering no grid (HNG), 0% (0/40). An overall 4 (groups) × 2 (solution/no solution) chi-square analysis was conducted. The chi square was significant, $\chi^2(3, N = 160) = 10.97$, $p < .05$, $\lambda = .05$. There was a difference in solution rates between the four groups.

Two follow-up tests explored the effect of the two manipulations. A 2 (training type) × 2 (solution/no solution) chi-square analysis was conducted to investigate the effect of the non-dot

[1] Thanks to Tom Ormerod for questioning why multiple non-dot turns would be more difficult to make than one.

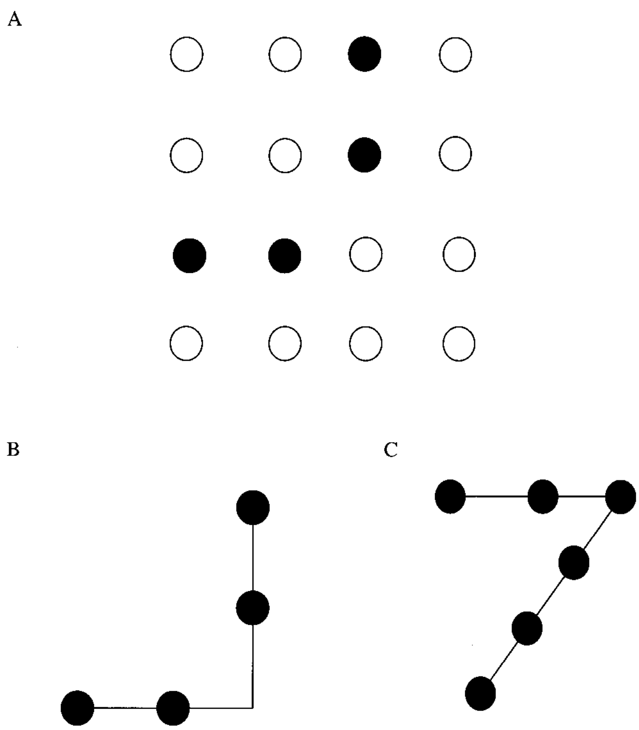


*Figure 3.* Examples of facilitating training on grid (non-dot turns; A), solution to facilitating training problem (no grid; B), solution to hindering (dot turns) training (C).

turns training. Participants in the facilitating training groups had a higher solution rate (12/80) than participants in the hindering training groups (1/80), $\chi^2(1, N = 160) = 10.13$, $p < .05$, $\lambda = .12$.

A separate 2 (grid/no grid) × 2 (solution/no solution) chi-square analysis was conducted to determine the overall effect of the grid manipulation. No difference was found between participants who received training on a grid (8/80) and participants who did not receive training on a grid (5/80), $\chi^2(1, N = 160) = .754$, $p > .05$, $\lambda = .03$.

### *Discussion*

Participants who learned to make non-dot turns were more likely to solve the nine-dot problem. This result provides support for the hypothesis that no prior experience or unhelpful prior experience is one source of difficulty in the nine-dot problem. As the multiple factors hypothesis predicts, the effect was quite small in magnitude, increasing the solution rate by a mere 15% for the grid groups and by 12.5% for the no-grid groups. The fact that training in making non-dot turns helped also supports the identification of non-dot turns as the key action in the nine-dot problem.

There was no support for the hypothesis that the figure–ground relationship influenced performance on the nine-dot problem. Participants who received their training on a grid were no more likely than participants in the no-grid group to succeed. It is possible that the figure–ground factor is not operating in this problem.

An alternative explanation is that the participants in the grid groups failed to transfer what they had learned from the training problems (all of which had a grid) to the no-grid presentation of the target problem. In a follow-up to Experiment 1, 10 participants solved 12 training exercises like those in Figure 3A and 3B. Each exercise was presented first on a page with a grid and then again on a second, plain page. The procedure was otherwise the same as for the facilitating-grid group. The result was that 30% (3/10) solved the nine-dot problem, which is nearly twice the solution rate for the facilitating-grid group (17.5%). The possibility that the figure–ground factor operates in the nine-dot problem should not be dismissed.

## Experiment 2

The goal state for the nine-dot problem is described to problem solvers abstractly (use four lines to connect all the dots) but not concretely, that is, the participants are not shown what the set of lines look like, once drawn. Greeno (1978), Simon (1973), and others have hypothesized that an ill-defined goal makes a problem more difficult. Not knowing which line configuration he or she is aiming for makes the search through the space of possible line configurations less selective. Experiment 2 tested this potential source of difficulty by teaching the participants the visual appearance of the goal configuration before they attempted the problem. Even if one knows which four lines are to be drawn, one still has to figure out in which sequence they are to be drawn (Burnham & Davis, 1969). Half the participants drew the relevant goal configuration, half learned to recognize it. The control groups drew and recognized an irrelevant figure.

### *Method*

#### *Participants and Design*

Participants were 134 undergraduates from UIC's participant pool. Fourteen participants were dropped from the final analysis because of prior knowledge of the nine-dot problem, which left 40 in each group. No demographic data were collected about the participants.

Experiment 2 utilized a 2 × 2 design. The independent variables were type of training (motor or visual) and training shape (relevant or irrelevant).

#### *Materials*

All materials were contained within booklets. The booklets for the motor training groups were 23 pages long, and the booklets for the visual training groups were 38 pages long. The booklets contained instructions, the training exercises, the target problem, and screening for familiarity with the nine-dot problem.

*Motor training.* The training exercises for the relevant-shape group consisted of numbered dots that, when connected, formed the arrowlike shape of the nine-dot problem (see Figure 4). Each training exercise was displayed on a separate sheet of paper. The dots were arranged in different orientations for each training exercise, such that the arrow shape was rotated to varying degrees. For the irrelevant-shape group, the numbered dots were arranged to form a geometric shape that did not correspond to the shape of the solution to the nine-dot problem (see Figure 4). In both groups, points at which a non-dot turn was needed were marked with open circles. There were 12 training exercises for the motor training groups.

*Visual training.* The training exercises for the visual training groups consisted of a set of target and distractor shapes (see Figure 5). In the relevant-shape groups, 12 target figures matched the solution shape of the nine-dot problem. In the irrelevant-shape groups, 12 target shapes matched a four-line geometric shape that did not correspond to the shape of the nine-dot problem. The 12 distractor shapes were the same for both groups,

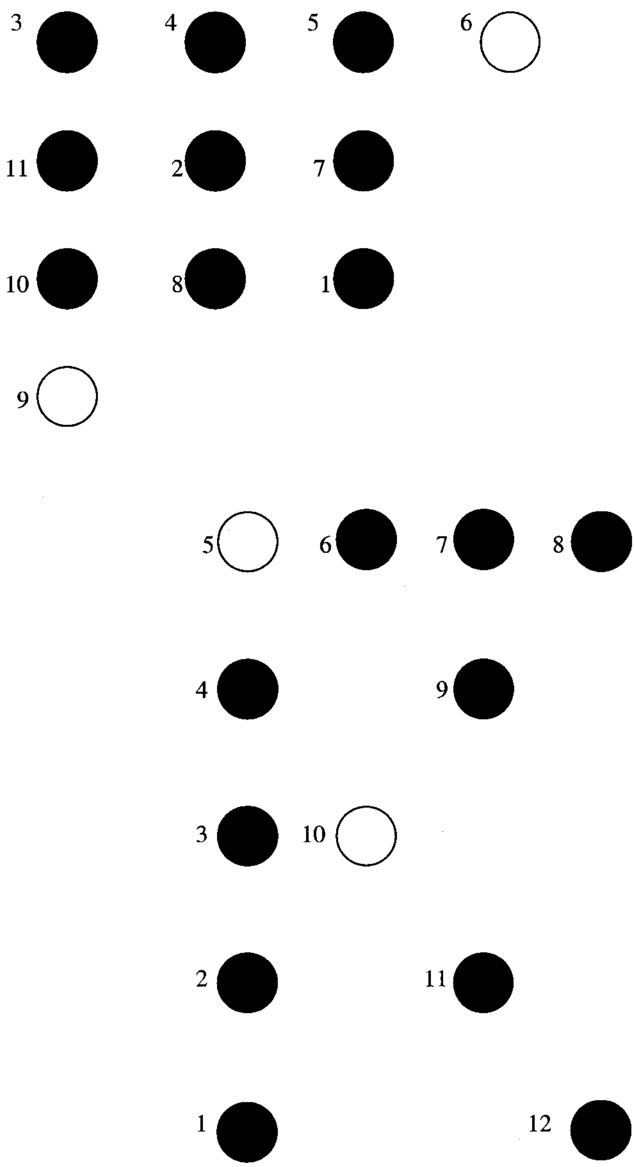

*Figure 4.* Motor training with relevant (top) and irrelevant (bottom) conditions.

and consisted of open and closed multiple line figures. In addition, there were three copies of a page of 13 shapes, 12 distractor and 1 target, in which the participants were required to recognize and circle the target shape.

## Procedure

*Motor training.* The 12 training exercises were presented in random order. The participants were instructed to connect the dots in the order indicated by the numbers next to the dots (see Figure 4). They were not told that the shape was the solution to a problem. The participants had 1 min to complete each exercise. After every four exercises (three times total), participants were instructed to draw the same shape on a blank piece of paper. Participants had 1 min to draw the shape.

*Visual training.* The experimenter showed the participants a 3 × 3 square of dots on a screen via an overhead projector and then placed a slide with a configuration of four lines on top of the dots. The presentation of the slides on the overhead lasted 30 s. For the relevant-shape group, the four lines formed the solution to the nine-dot problem (see Figure 5A). The participants were not told that the line configuration was the solution to a problem, but they were instructed to remember the shape for use during some training exercises. For the participants in the irrelevant shape group, the line configuration formed an irrelevant shape (see Figure 5A). The participants were not given any hints as to possible orders in which the four lines in the configuration might be drawn.

The training exercise for the visual training groups was a judgment task in which participants had to determine if a shape that they were viewing was the same as or different from a shape they learned prior to the judgment task. For the relevant-shape group, the participants compared new shapes to the shape of the solution to the nine-dot problem. In the irrelevant-shape group, the participants compared new shapes to an irrelevant shape (see Figure 5 for target and distractor shapes). There were 24 (12 target and 12 distractor) judgments. For each sheet, the participants were asked to judge if the shape was the same as or different than the shape the experimenter showed them initially. After making each judgment, participants rated how confident they were of their judgment on a scale of 1 (*not confident at all*) to 5 (*totally confident*). After every eight judgments (three times total), participants were given a shape recognition task. They were presented with a sheet that had 13 shapes on it, and they had 1 min to identify the shape they had been shown at the beginning. For participants in the relevant shape group, the multiple-configuration sheets included the solution to the nine-dot problem. For the irrelevant shape group, it was replaced by the irrelevant shape (see Figure 5A).

*Target problem.* After completing the training exercises, all participants had 4 min to complete the nine-dot problem, using the same two-sheet presentation format as in Experiment 1. At the end of the experiment,

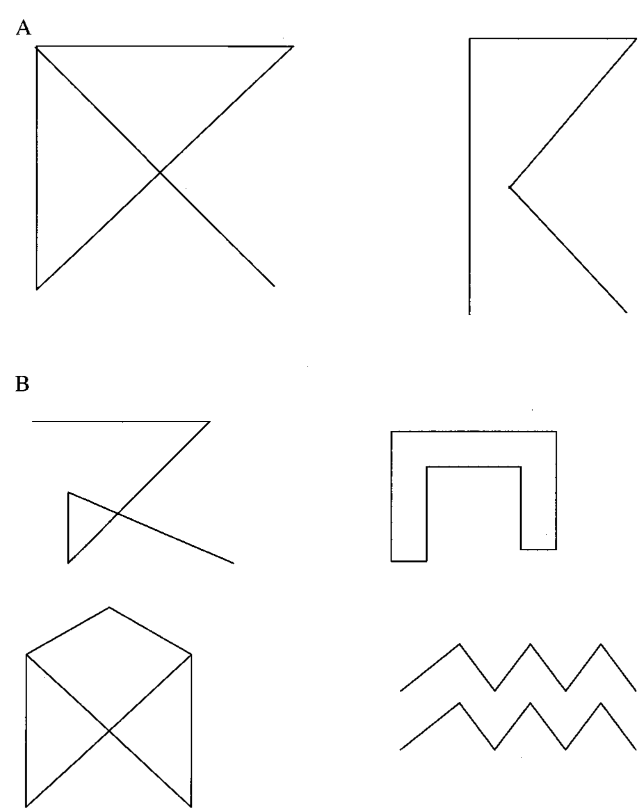


*Figure 5.* Visual training with relevant and irrelevant shapes (A) and examples of distractor shapes (B).

participants were screened for prior knowledge of the nine-dot problem, as in Experiment 1.

### *Results*

Solution rates for the nine-dot problem across groups are as follows: motor training, relevant shape, 10% (3/30); motor training, irrelevant shape, 0% (0/30); visual training, relevant shape, 13.3% (4/30); visual training, irrelevant shape, 3.3% (1/30). A 4 (groups) $\times$ 2 (solution/no solution) chi-square analysis did not show any overall differences between the groups, $\chi^2(3, N = 120) = 5.36$, $p > .05$, $\lambda = .04$.

A 2 (relevant vs. irrelevant) $\times$ 2 (solution/no solution) chi-square analysis was conducted to determine the effect of shape. Participants who received a relevant shape (7/60) performed better than participants who received an irrelevant shape (1/60), $\chi^2(1, N = 120) = 4.82$, $p < .05$, $\lambda = .03$.

A second 2 (motor vs. visual) $\times$ 2 (solution/no solution) chi-square analysis was conducted to determine the overall effect of type of training. There was no significant difference between solution rates for the motor groups (3/60) and the visual groups (5/60), $\chi^2(1, N = 120) = .54$, $p > .05$, $\lambda = .09$.

### *Discussion*

Participants who learned the shape of the relevant line configuration performed better on the nine-dot problem than participants who learned an irrelevant shape. This supports the hypothesis that knowing the target shape that one is aiming for helps reduce (a) the uncertainty in the search through the space of possible line configurations and (b) the amount of processing the problem requires. This type of training was different in character from the non-dot turns training in Experiment 1, but it also influenced solution rate. As the multiple-factors hypothesis predicts, the effect was small in magnitude. The solution rate was only raised by 10%. It did not matter whether the desired goal configuration was learned via the motor or the visual modalities.

## Experiment 3

Experiments 1 and 2 showed that prior experience of making non-dot turns and advance knowledge of the goal configuration improve the solution rate on the nine-dot problem. These two facilitating procedures are different in character and have different rationale (overcome prior knowledge, cut back search), yet both produce a positive effect. Both effects were small in magnitude.

Experiment 3 extends these results by testing the hypothesis that if we combine multiple components into a single facilitation procedure we should observe a larger increase in solution rate. In Experiment 3, participants were trained in how to make non-dot turns, and they were taught the visual appearance of the desired line configuration (but not how to draw it). In addition, the facilitation procedure included the two perceptual sources of difficulty that we hypothesize are operating in the nine-dot problem. The classical good Gestalt factor (people tend to assume that all lines are to be drawn inside the square formed by the dots) was counteracted with the help of target problems in which the dots do not form a square or other good Gestalt. We alleviated the figure–ground factor (people tend to assume that lines cannot extend into, or turn on, the white background) by presenting training problems on a full page grid and by providing an opportunity for the participants to transfer to a similar problem on a plain background before encountering the target problem.

In addition, Experiment 3 further tested the hypothesis that a non-dot turn is the key action that is difficult to produce. The target problems varied in the number of non-dot turns required for solution. We expected that the larger the number of non-dot turns, the more difficult the problem. This prediction is based on how the perceptual, knowledge, and process factors interact to make turning on a non-dot point an unlikely action. Even when a participant has made one non-dot turn, and therefore has overcome any biases of prior knowledge, he or she still needs to overcome the perceptual and process factors involved in making non-dot turns. The participant will need to view all white space surrounding the problem as an acceptable drawing area and, more important, determine the correct location at which to make the next non-dot turn.

The 11-dot variant (cf. Burnham & Davis, 1969; MacGregor et al., 2001; see Figure 6D) requires no non-dot turns and should therefore be easiest. The 10-dot problem (see Figure 6A) has an additional dot at the lower left-hand corner and hence requires one turn on a non-dot point. It should therefore be more difficult than the 11-dot variant.

The displaced nine-dot problem (see Figure 6B), in which the top line of dots in the nine-dot problem is shifted to the right by one dot, also requires two non-dot turns. Another variant moves the dot in the upper left hand corner to the bottom right hand corner, as an extension to the right of the figure (see Figure 6C). This variant requires three non-dot turns and is therefore called the three-turn problem. This should be the most difficult of the problem variants.

The nine-dot problem has two non-dot turns and should therefore be as difficult as the displaced non-dot problem. However,

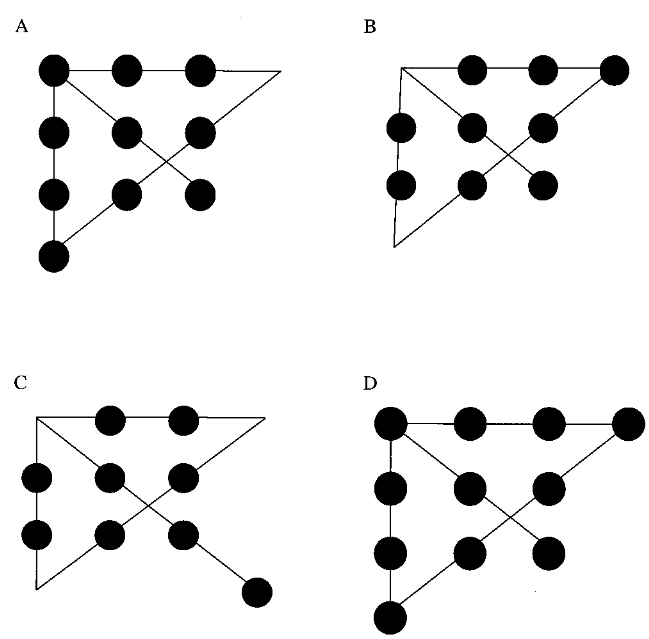


*Figure 6.* The 10-dot (A), displaced nine-dot (B), three-turn (C), and 11-dot (D; Burnham & Davis, 1969) problems.

two additional factors moderate this prediction. First, the nine-dot problem should have an even lower solution rate because the good Gestalt factor traditionally emphasized by the Gestalt psychologists is present in the nine-dot problem but not in the displaced nine-dot problem. Second, the three-turn problem should have a solution rate that is lower than the nine-dot problem. However, due to the very low solution rate of the nine-dot problem, it is not expected that the three-turn problem will have a significantly lower solution rate than the nine-dot problem in the control condition.

## *Method*

### *Participants and Design*

Participants were 332 undergraduates from UIC's participant pool. Thirty-two participants (15 from the control group and 17 from the experimental group) were dropped from the final analysis because of prior knowledge of the nine-dot problem, which left 300 participants in the analysis. No demographic data were collected about the participants.

The design of Experiment 3 was a 2 $\times$ 5 design with two types of training (experimental and control) and five target problems: 11-dot, 10-dot, displaced nine-dot, three-turn, and nine-dot.

### *Materials*

The booklet for the control groups was five pages long and contained instructions, a target problem, and screening for familiarity with the nine-dot problem. The booklets for the experimental groups were 33 pages long and contained basic instructions, training exercises, a target problem, and screening for familiarity with the nine-dot problem.

The non-dot training exercises in Experiment 3 were taken from Experiment 1 (see Figure 3A and 3B). There were 12 exercises that involved connecting dots. Another section of the training involved making perceptual judgments, as in Experiment 2 (see Figure 5 for the target and distractor shapes). The control groups did not receive any training.

The target problems were the 11-dot (Burnham & Davis, 1969), the 10-dot, the displaced nine-dot, the three-turn, and the traditional nine-dot (see Figure 6). Each participant solved one target problem. The target problems were presented in the same two-page format as in Experiments 1 and 2.

### *Procedure*

The following description details the procedures followed by the experimental group. Participants in the control group attempted a target problem without any training.

At the beginning of the experiment, the experimenter told the participants that they would be learning several different pieces of knowledge that they would need to complete a problem at the end of the experiment. Using an overhead projector, the experimenter showed the participants the shape of the solution to the nine-dot problem and described it as having three lines that formed a triangle and a fourth line that cut the triangle in half (see Figure 5). Participants were instructed to remember the shape because they would need it for the next exercise and for a problem at the end of the experiment. They were not shown how the four lines might be drawn.

The participants then worked on a perceptual judgment task that followed the same visual training procedure as in Experiment 2, with the exception that participants were not shown the pattern of dots that make up the nine-dot problem and did not complete any shape recognition tasks. Participants made 24 (12 target and 12 distractor) judgments and confidence ratings. After every judgment, the experimenter showed the participants the correct answer for that judgment and reminded the participants about the important features of the shape (a triangle cut in half by a fourth line). After completing the visual training, participants were once again told that they should remember the shape for use later in the experiment.

In the second part of the training, participants were instructed to connect dots by using a certain number of straight lines, analogous to the training in Experiment 1. The figures were presented in pairs, as in the follow-up to Experiment 1, such that a figure appeared on a grid of unfilled circles and subsequently appeared on a plain page. There were six pairs or 12 problems total. The experimenter said that there would sometimes be unfilled circles on the page and that the participants should use the unfilled circles to connect the dots. Participants were also encouraged to remember where in the figure the unfilled circles were used because that would be important in solving a problem at the end of the experiment. After every two exercises, the experimenter showed the participants the correct answer for the previous pair of problems and pointed out that for the first problem of the pair unfilled circles were used and for the second problem of the pair the participants had drawn lines in blank space.

After the completion of the second part of the training, the experimenter reviewed the purpose of learning to connect dots and emphasized the importance of non-dot turns for completing the final problem. The experimenter then introduced the target problem by saying that the participants should remember what they learned about the shape of the problem and how to make non-dot turns when working on the target problem. The experimenter also suggested that participants could draw unfilled circles around the dots if they found it difficult to turn in empty space. The participants had 4 min to attempt the target problem. The target problems were presented, as in Experiments 1 and 2, on two pages of the experimental booklet.

## *Results*

Solution rates for each target problem across groups are presented in Table 2. An overall 10 $\times$ 2 chi-square analysis was conducted. The chi square was significant, $\chi^2(9, N = 300) = 122.68$, $p < .05$, $\lambda = .43$. There were differences in solution rates

Table 2
*Experiment 3: Group Solution Frequencies for the Target Problem*

| | | Solution rate | | |
|---|---|---|---|---|
| Problem | No. of non-dot turns | Experimental | Control | Total |
| 11-dot | 0 | 29 (97) | 16 (53) | 60 (75) |
| 10-dot | 1 | 24 (80) | 2 (7) | 26 (43) |
| Displaced 9-dot | 2 | 15 (50) | 1 (3) | 16 (27) |
| 9-dot | 2 | 12 (40) | 2 (7) | 14 (23) |
| 3-turn | 3 | 9 (30) | 3 (10) | 12 (20) |

*Note.* Each group had 30 participants. Percentages are shown in parentheses.

between the 10 groups. Follow-up tests were done between the collapsed control versus experimental groups, between the problem types within the control group, and between the problem types within the experimental group.

### Effect of Training

A 2 × 2 chi-square analysis was conducted to determine the overall effect of training. Participants in the training groups performed better at 59.3% (89/150) than participants in the control groups at 16% (24/150), $\chi^2(1, N = 300) = 59.98$, $p < .05$, $\lambda = .25$. The training had an effect, and the effect was larger, 43.3%, than in Experiments 1 and 2.

### Effect of Number of Non-Dot Turns

A 5 × 2 chi-square analysis showed overall differences between the target problems, $\chi^2(4, N = 300) = 52.76$, $p < .05$, $\lambda = .27$. There were differences in solution rate between the five problems. Recall that the five target problems differ in the number of non-dot turns required for solution.

A 5 × 2 chi-square analysis was conducted to determine the effect of target problem within the control group. There was a significant difference between the solution rates for the problem types, $\chi^2(4, N = 150) = 39.39$, $p < .05$, $\lambda = .08$. The standardized residuals for each cell were examined. The participants who had solved the 11-dot problem (53.3% or 16/30) caused the highest standardized residual, 5.1; therefore, this cell made the greatest contribution to the chi square. Recall that the 11-dot problem required no non-dot turn. When the 11-dot problem was removed from the analysis, differences between the problem types were no longer significant, $\chi^2(3, N = 120) = 1.07$, $p > .05$, $\lambda = .02$. The solution rates for those problems that required one or more non-dot turns did not differ. As shown in Figure 7, the solution rates for those problems were all at floor.

A 5 × 2 chi-square analysis was conducted to determine the overall effect of target problem within the experimental group. There was a significant difference between solution rates, $\chi^2(4, N = 150) = 39.07$, $p < .05$, $\lambda = .30$. The standardized residuals for each cell were examined. The participant who had not solved the 11-dot problem (3.3% or 1/30) caused the highest standardized residual, −3.2; therefore, this cell made the greatest contribution to the chi square. When the 11-dot problem was removed from the analysis, differences between the four remaining problem types were still significant, $\chi^2(3, N = 120) = 16.80$, $p < .05$, $\lambda = .22$. However, when both the 11-dot problem (no non-dot turn) and the 10-dot problem (one non-dot turn) were removed from the analysis, there was no significant difference between the remaining three problem types, $\chi^2(2, N = 90) = 2.50$, $p > .05$, $\lambda = .06$. That is, the problems that required two non-dot turns were not measurably easier than the three-turn problem. The differences between the groups can be seen in Figure 7.

Another way to consider the effect of the number of non-dot turns is to determine the probability of making a non-dot turn. For both the experimental and control groups, we calculated the percentage of participants who made non-dot turns versus correct non-dot turns in the displaced nine-dot, traditional nine-dot, and three-turn problems, each of which requires more than one non-dot turn. In the control group, 24% (22/90) of the participants made

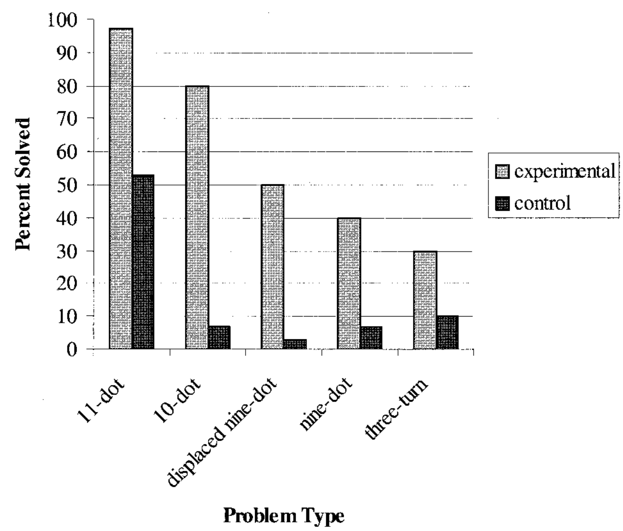


*Figure 7.* Experiment 3: Solution rates for each target problem.

one non-dot turn. Of those that made one non-dot turn, 77% (17/22) made two non-dot turns; of those who made two non-dot turns, 67% (4/6 on the three-turn problem) made three non-dot turns. However, only 11% (10/90) made one correct non-dot turn. Of those that made one correct non-dot turn, 30% (3/10) made two correct non-dot turns; of the latter, 50% (3/6 on the three turn problem) made three correct non-dot turns.

In the experimental group, 76% (68/90) made one non-dot turn. Of those that made one non-dot turn, 75% (51/68) made two non-dot turns; of the latter, 75% (12/16 on the three-turn problem) made three non-dot turns. However, 70% (63/90) made one correct non-dot turn. Of those that made one correct non-dot turn, 54% (34/63) made two correct non-dot turns; of the latter, 90% (9/10 on the three-turn problem) made three correct non-dot turns.

### Effect of the Gestalt Factor

The nine-dot problem forms a good Gestalt, but the dot groups for the other four target problems do not. To assess this classical perceptual factor, we compared the number of participants who drew lines outside the figure formed by the dots in the nine-dot problem with the number of participants who did so in the other target problems. Participants who solved the problem and those who did not were both included in this analysis. In the control groups, 25.8% (31/120) of the participants drew lines outside the dots. The differences between the problem types within the control group were not significant, $\chi^2(3, N = 120) = 5.35$, $p > .05$, $\lambda = .05$.

In the experimental groups, 82.5% (99/120) of the participants drew lines outside the dots (see Table 3). Measured by the power to overcome the Gestalt factor, the effect of the training procedure was 56.7%. The fact that the tendency to draw such lines is affected by training shows that the good Gestalt factor is indeed operating in this domain.

Furthermore, we can compare the tendency to draw lines outside the dots across the problem types. In three of the problem types (10-dot, displaced nine-dot, and three-turn problem), participants were equally likely to draw lines outside the dots, $\chi^2(2, N = 90) =$

Table 3
*Experiment 3: Number of Lines Drawn Outside the Dots for Each Group and Problem Type*

| Problem type | No. non-dot turns | Control | | Experimental | |
|---|---|---|---|---|---|
| | | Outside | Inside | Outside | Inside |
| 10-dot | 1 | 5 | 25 | 26 | 4 |
| Displaced 9-dot | 2 | 10 | 20 | 28 | 2 |
| 9-dot | 2 | 5 | 25 | 18 | 12 |
| 3-turn | 3 | 11 | 19 | 27 | 3 |

*Note.* The number of lines drawn outside was not counted for the 11-dot problem because the problem does not require any non-dot turns.



.74, $p > .05$, $\lambda = .03$. But when participants received the nine-dot problem, they were less likely to draw lines outside the dots, thus providing evidence for the good Gestalt of the square. This effect is particularly striking when the nine-dot problem is compared to the displaced nine-dot, $\chi^2(1, N = 60) = 9.31$, $p < .05$, $\lambda = .23$. Although the solution rate for these two problems did not differ, the displaced nine-dot problem led to a greater incidence of drawing lines outside the dots than the traditional nine-dot problem.

## *Discussion*

The combined training procedure, designed to alleviate a perceptual factor, a knowledge factor, and a process factor, was effective and raised the solution rate to 40%, a considerably larger effect than those observed in Experiments 1 and 2. If we estimate the effects for the relevant factors to be 10–15% (in our subject population), then an additive prediction would place the expected effect of the three combined manipulations within the range 30–45%, which is what we observed. This result is consistent with the hypothesis of multiple sources of difficulty.

The hypothesis that non-dot turns are difficult is supported by the fact that only 24% of the participants in the control group made one or more non-dot turns. In addition, the solution rate is strongly influenced by the number of non-dot turns. Within both the control and the experimental groups, all problems that require at least one non-dot turn were more difficult than the one problem that requires none. Furthermore, within the experimental groups, the problems that required two or more non-dots were more difficult than the problem that required a single such turn. Participants in the experimental group were more likely to make both a first non-dot turn and a first correct non-dot turn than participants in the control group because of the training they had received in making non-dot turns. However, the participants in the experimental group, despite extensive training, had only a 50% chance of making a correct second non-dot turn, which shows the pervasiveness of the process factor. These observations are consistent with the idea that making non-dot turns is the key action in these problems.

## General Discussion

Insight problems are interesting in part because they are simple, yet capable of keeping adults puzzled for minutes, sometimes hours. Single-factor theories of difficulty such as functional fixedness (Duncker, 1945) and constraint relaxation (Ohlsson, 1992; Knoblich et al., 1999) predict that hints and training that counteract that difficulty should raise the solution rate to 100%. With the exception of studies in which the participants were given half the solution (Weisberg & Alba, 1981), empirical outcomes are not consistent with this prediction. We hypothesize that insight problems are difficult because the key behavior needed for solution tends to be suppressed by multiple, accidentally converging factors related to perceptual factors (e.g., good Gestalt, figure–ground), prior knowledge and experience, and processing demands (e.g., amount of lookahead). This hypothesis predicts that qualitatively different single-factor manipulations might cause significant but small increases in solution rate, as we observed in Experiments 1 and 2. It also predicts that combined manipulations that address more than one difficulty will produce larger increases in solution rate, as we observed in Experiment 3.

The studies reported here are limited in at least two ways. First, even with the combined training in Experiment 3, not every participant solved the nine-dot problem. There remains the possibility that we have not identified all the difficulty factors operating even in this simple problem. To clinch the case for the multiple-factors perspective, we need to demonstrate cumulative effects of combining an increasing number of training components, with the strongest training producing a solution rate of 100%.

The second limitation is that these studies only apply the multiple-factors perspective to a single task. However, it does have plausibility with respect to other insight problems. For example, the hat rack problem (Maier, 1933) requires that two boards are wedged between floor and ceiling with the help of a clamp and that the handle of the clamp be used as a hanger. Perceptual difficulty factors might in this case include the fact that the ceiling of a room is not normally in our perceptual field unless we have a reason to look up and biasing knowledge factors might include prior experience of achieving stability via support from below rather than via wedging, as well as the prior habit of seeing a handle on a clamp as merely a handle. In contrast, processing factors might dominate in the necklace problem (Wickelgren, 1974). There are many combinations of links that one might try to open, and the need to keep track of the cost of each elementary operation might pose demands on working memory. Although conjectural at this point, these analyses illustrate the idea that each insight problem might involve a unique combination of perceptual, knowledge, and process factors that accidentally combine to suppress the key action or actions in the solution.

This raises the question of how widely the principle of multiple difficulty factors might apply. It is possible that creative accomplishments outside the laboratory, for example, technological inventions that seem obvious in hindsight, were so difficult to think of because the key moves were suppressed by multiple difficulty factors.

## References

Ansburg, P. I., & Dominowski, R. L. (2000). Promoting insightful problem solving. *Journal of Creative Behavior, 34*(1), 30–60.

Beeman, M. J., Haberman, J., & Bowden, E. M. (2002, November). *fMRI signal at the moment of insight, during insight-like problems.* Paper presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Birch, H. G., & Rabinowitz, H. S. (1951). The negative effect of previous experience on productive thinking. *Journal of Experimental Psychology, 41,* 121–125.

Bowden, E. M., & Beeman, M. J. (1998). Getting the right idea: Semantic activation in the right hemisphere may help solve insight problems. *Psychological Science, 9*(6), 435–440.

Burnham, C. A., & Davis, K. G. (1969). The nine-dot problem: Beyond perceptual organization. *Psychonomic Science, 17*(6), 321–323.

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*(3), 903–919.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58*(5, Whole No. 270).

Gick, M. L. (1989). Two functions of diagrams in problem solving by analogy. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 215–231). Amsterdam: North Holland (Elsevier Science Publishers).

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.

Greeno, J. G. (1978). Natures of problem-solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 5. Human information processing* (pp. 239–270). Hillsdale, NJ: Erlbaum.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374–419.

Kershaw, T. C., & Ohlsson, S. (2001). Training for insight: The case of the nine-dot problem. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 489–493). Mahwah, NJ: Erlbaum.

Kershaw, T. C., Ohlsson, S., & Coyne, C. (2003). The fallacy of single-source explanations: The multiple difficulties of the nine-dot problem. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* [CD-ROM]. Cognitive Science Society.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1534–1555.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory and Cognition, 29*(7), 1000–1009.

Köhler, W. (1971/ orig. 1930). Human perception. In M. Henle (Ed.), *The selected papers of Wolfgang Köhler* (pp. 142–167). New York: Liveright.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*(2), 143–183.

Lavric, A., Forstmeier, S., & Rippon, G. (2000). Differences in working memory involvement in analytical and creative tasks: An ERP study. *Neuroreport, 11*(8), 1613–1618.

Lung, C. T., & Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 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, and Cognition, 27*(1), 176–201.

Maier, N. R. F. (1930). Reasoning in humans: I. On direction. *Journal of Comparative Psychology, 10,* 115–143.

Maier, N. R. F. (1933). An aspect of human reasoning. *British Journal of Psychology, 12,* 144–155.

Maier, N. R. F., & Casselman, G. G. (1970). Locating the difficulty in insight problems: Individual and sex differences. *Psychological Reports, 26,* 103–117.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(2), 288–294.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and non-insight problem solving. *Memory and Cognition, 15*(3), 238–246.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S. (1984a). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25*(1), 65–78.

Ohlsson, S. (1984b). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25*(2), 117–129.

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

Ormerod, T. C, MacGregor, J. N., & Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 791–799.

Scheerer, M. (1963) Problem solving. *Scientific American, 208*(4), 118–128.

Schooler, J. W., & Melcher, J. M. (1995). The ineffability of insight. In S. M. Smith & T. B. Ward (Eds.), *The creative cognition approach* (pp. 97–133). Cambridge, MA: MIT Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166–183.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4,* 181–201.

Smith, S. M., Ward, T. B., & Finke, R. A. (1995). *The creative cognition approach.* Cambridge, MA: MIT Press.

Sternberg, R. J., & Davidson, J. E. (1995). *The nature of insight.* Cambridge, MA: MIT Press.

Weisberg, R. (1986). *Creativity: Genius and other myths.* New York: 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*(2), 169–192.

Wickelgren, W. A. (1974). *How to solve problems.* San Francisco: Freeman.

Received March 12, 2003
Revision received July 18, 2003
Accepted July 21, 2003 ■