THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2001, *54A* (3), 903–919

When insight just won't come: The failure of visual cues in the nine-dot problem

Edward P. Chronicle and Thomas C. Ormerod
Department of Psychology, Lancaster University, Lancaster, UK
James N. MacGregor
School of Public Administration, University of Victoria, Victoria, Canada

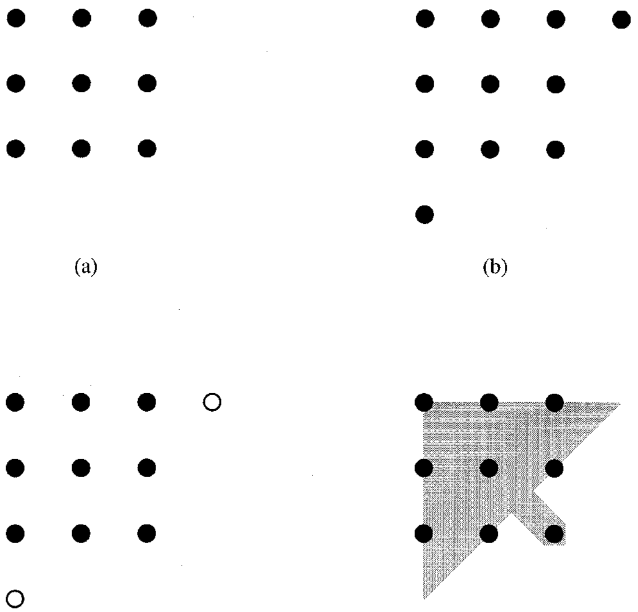
The nine-dot problem is a classic in the field of human problem solving. Cognitive accounts of the problem's difficulty have been criticized on the grounds that the experimental methods on which they rely for support involve a qualitative change to the task requirements of the problem. The three experiments reported here utilize visual and visual–procedural hints to examine the notion that its difficulty is rooted in a mismatch between problem shape and solution shape. Experiment 1 demonstrated that a perceptual cue to the shape of the solution in the form of shading gave rise to only minimal improvements in performance; an additional hint about the relevance of the shading gave rise to modest, but not statistically significant, improvements. Experiment 2 replicated these findings against an additional control condition in which a solely verbal hint to violate the perceptual boundary of the problem shape was given. Furthermore, when both the verbal and visual hints were provided, performance improved only slightly. Experiment 3 provided participants with experience in producing the shape of the correct solution in problem variants closely related to the nine-dot problem. Performance on the transfer task, the basic nine-dot problem, remained at floor, however. These data suggest that visual constraint relaxation is unlikely to be the sole process by which the insight required to find a solution is achieved. The results are interpreted in terms of a previously proposed computational model of performance.

The nine-dot problem is a classic problem in the psychology of thinking, and it provides a powerful demonstration of human failure to solve what appears to be a straightforward spatial problem with a simple task description. The problem is illustrated in Figure 1a. The task is to connect all nine dots with four straight lines, drawn without lifting pencil from paper or retracing a line. The great majority of published accounts of performance on the problem report that 0% of participants are able to solve it, despite experimental manipulations that utilize a large number of participants (Weisberg & Alba, 1981, Experiment 1b: approximately 50 participants), allow a good deal of time for a solution to be attempted (Burnham & Davis, 1969:

Requests for reprints should be sent to Tom Ormerod, Department of Psychology, Lancaster University, Lancaster, LA1 4YF, UK. Email: T.Ormerod@lancaster.ac.uk

Preliminary reports of the experiments described in this paper were given at a meeting of the Experimental Psychology Society, Oxford, 18–19 March, 1999, and Mind III: Spatial Cognition, Dublin, 17–19 August, 1998.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980042000471



(a) (b) (c) (d)

Figure 1. (a) The standard nine-dot problem; (b) the eleven-dot version used by Burnham and Davis (1969); (c) the eleven-dot variant used by Ormerod et al. (1997); and (d) the shading hint used in Experiments 1 and 2. In all cases, the requirement is to draw four connected straight lines through all the dots, without lifting pen from paper, and without retracing.

10 min per attempt), or give participants many separate attempts (Weisberg & Alba, 1981, Experiment 1a: 100 attempts). The highest solution rate reported for naïve participants was 9.4% (Lung & Dominowski, 1985). A popular explanation for the problem's difficulty is that perception of the dot array is governed by the Gestalt law of Pragnanz, causing the mental representation of the dots as a square (Scheerer, 1963). Participants are unable to solve the problem because they fixate upon the square, and consequently fail to consider solution options that do not fall within the boundary of this figure. This account persists in textbooks of psychology (e.g., Groome, 1999; Wade & Tavris, 1990) despite the existence of alternative accounts that propose cognitive rather than perceptual factors as the source of the problem's difficulty (e.g., Lung and Dominowski, 1985; Weisberg & Alba, 1981).

Typically, cognitive accounts seek support from some form of instructional or training manipulation that facilitates, or conversely fails to facilitate, problem-solving performance in a way that cannot, it is argued, be accounted for within a Gestalt tradition. For example, Burnham and Davis (1969) report a number of instructional manipulations that led to only small amounts of facilitation. The most important manipulation was to instruct participants that they could go outside the square formed by the dots, which was intended to override the set formed by perceptual grouping of the dots. Participants receiving this instruction did seek solutions by drawing lines outside the dot array perimeter, unlike those receiving no instruction, but only 4 out of 15 participants successfully solved the problem. Burnham and Davis suggest that the failure of this manipulation to remove all aspects of the problem's difficulty argues against an account in which problem difficulty is caused by the imposition of a rigid perceptual organization on the dot array. Similar results are reported by Weisberg and Alba (1981, Experiment 1), who instructed participants that they had exhausted all the solution possibilities that lay within the square (we subsequently refer to this as the "exhausted square" hint), and that they would have to go outside the square in order to solve the problem. They found that only 3 out of 15 participants given this hint solved it within 10 attempts.

The conclusion drawn by Burnham and Davis (1969) and by Weisberg and Alba (1981) was that the failure of participants to solve the problem when instructed to explore solution options outside the square falsifies a Gestalt account of difficulty, as this manipulation explicitly overrides the self-imposed perceptual organization of the problem array. This argument seems to us to be conceptually flawed. The operation of the Gestalt law of Pragnanz (e.g., as implemented in Marr's, 1982, theory of the full primal sketch) is at an automatic and pre-conscious level and cannot be overridden by conscious control. Thus, although participants may be instructed to act contrary to their perceptual processing of the problem array, perceptual organization takes place regardless and may well continue to exert an influence over problem-solving performance in opposition to overt instruction. So, far from demonstrating that an explicit dismissal of the "square" interpretation fails to facilitate performance, the instructional manipulation may simply be unsuccessful in overriding the organization and subsequent influence of the problem's perceptual characteristics.

Weisberg and Alba (1981) also report data from their first experiment in which performance was strongly facilitated by presenting parts of the solution along with the "exhausted square" hint. In one condition one line of the solution was already drawn (a straight diagonal line extending from a point outside of the square and passing through the bottom right-hand dot and extending to the top left dot), and 9 out of 13 participants solved the problem, taking an average of five attempts to do so. In a second condition participants were provided with the first two lines of a correct solution already drawn, the first being the diagonal line described above, the second being a horizontal line starting at the end of the diagonal line, passing through the dots on the top row, and extending to a non-dot point one unit to the right. All 17 participants in this condition solved the problem, generally on their first or second attempt. Weisberg and Alba suggest that the absence of complete facilitation with the one-line group indicates that the source of difficulty in the problem is not one of fixation upon the square figure. Although the facilitation shown by adding lines to the problem representation or by instructing problem-solvers on the characteristics of the required solution is undeniable, it is hardly surprising. Including in the instructions a component of the solution or a procedure for constructing the whole solution is surely bound to improve performance. Where we differ from Weisberg and Alba is in the interpretation of their findings. In the case of the nine-dot problem, the addition of solution components not only changes the visual display of the problem, but also qualitatively changes the performance options available to participants, effectively restricting them to only two starting points (the ends of the drawn lines). Facilitation is cognitively mediated in these experiments, but this may be the case only because the task faced by participants is changed qualitatively, by restricting the space of possible solutions and then making a search of the restricted space a task requirement. In other words, these manipulations show only that if you tell participants how to solve the problems, then they can solve them: They do not properly test the influence of perceptual organization upon problem-solving performance.

More compelling evidence for a cognitive basis to the nine-dot problem's difficulty comes from experiments in which facilitation emerges through transfer of experience from a training task. Weisberg and Alba (1981, Experiment 2) found that nearly half the participants who were trained on problems whose solutions required lines extending beyond the boundary of the problem array solved the nine-dot problem, whereas no participant trained on problems whose solutions remained within the problem array was able to solve the nine-dot problem. They interpret these results as demonstrating that performance on the nine-dot problem reflects the transfer of task-specific experience to the problem. People, they argue, generally encounter dot arrays such as the nine-dot problem in the context of joining-the-dots tasks, and therefore consider only the space of dot-to-dot lines in developing their solutions. Presenting lines that do not end on dots and training in solutions that have lines that do not terminate in dots create relevant experience that can be applied in the solution of the problem.

A similar training manipulation is reported by Lung and Dominowski (1985). They found that the most powerful facilitation is gained through a combination of training and instruction. They argue that it is not the retrieval of specific experience that blocks performance on the nine-dot problem. Instead, they suggest that it is the addition of a constraint that all lines must begin and end on dots that makes the problem difficult, as the solution requires lines to turn on a position where there is no dot to anchor them. This is similar to Weisberg and Alba's (1981) account, as the source of problem difficulty is proposed to be the failure to consider non-dot points in solution attempts.

The instructional and training manipulations reported in the literature do undoubtedly reduce problem difficulty, but they are all open to the same criticism that they present participants with components of the solution. One might argue that the most surprising aspect of these attempts at facilitation is that they are not perfect, suggesting that there must be some other aspect of the problem that determines difficulty beyond the conscious cognitive strategies adopted by participants. Interestingly, the strongest facilitation in nine-dot performance comes from a presentational manipulation. Burnham and Davis (1969) report that between 80% and 95% of participants were able to solve an eleven-dot variant of the problem (illustrated in Figure 1b) within 10 min. They interpret this manipulation as providing "a direction and rationale for drawing the necessary four lines for the solution"—that is, offering a template for the solution pattern. Having dismissed the Gestalt account on the basis of their "hint" data, they do not examine the possibility that the eleven-dot problem is facilitative precisely because the figure presented by the problem matches the figure required by the solution. However, their data are compatible with this problem–solution perceptual mapping account, and also with the dot-to-dot account of Weisberg and Alba (1981) and the dot-turning-points account of Lung and Dominowski (1985).

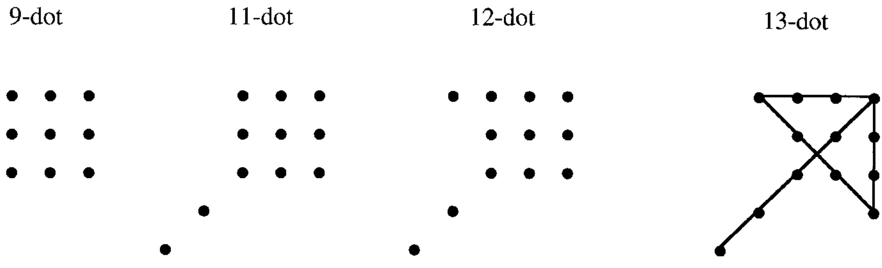
Previously, we proposed that problem difficulty arises because of figural factors, specifically because of the problem of mapping the arrow-like shape of the correct solution on to the Gestalt (a square) of the standard nine-dot problem (Ormerod, Chronicle, & MacGregor, 1997). We refer to this as the problem–solution perceptual mapping account. The eleven-dot problem of Burnham and Davis (1969) is easy to solve because the problem and solution both form arrow shapes. As a test of this hypothesis a revised version of the eleven-dot problem, shown in Figure 1c, was given to participants (Ormerod et al., 1997, Experiment 2). In this problem, dots that are additional to the standard nine-dot problem (the top right and bottom left dots) were left unfilled, and participants were instructed to solve the standard nine-dot problem using these additional dots as guides to the solution. Thus the availability of the extra dots for line extension and turning points was maintained but the perceptual salience of the arrow shape of the solution was reduced. After five trials, only 33% of participants solved this problem, compared with between 80 and 95% in Burnham and Davis' studies. An apparently minor perceptual change in the problem representation, in which the standard nine-dot problem was made distinct within the arrow-like eleven-dot problem, inhibited performance. We argued that the drop in solution rates could not be accounted for by either of the cognitive accounts of problem difficulty: Rather, the availability of solution shape information was the critical determinant of success. The provision of additional dots as guides and the instruction for participants to use them in constructing their solution effectively reduce the problem to one of drawing dot-to-dot lines as specified in Weisberg and Alba's (1981) account and remove the non-dot turning-points implicated in Lung and Dominowski's (1985) account. However, like some of the studies criticized earlier, our experiment (Ormerod et al., 1997) effectively changed the task demands of the problem: It may be, for example, that participants did not fully understand the instruction to utilize the unfilled dots in solving the nine-dot problem. To provide a proper test of previous accounts against our problem–solution perceptual mapping account requires interventions that do not qualitatively change the task demands of the problem itself.

More recently we proposed a computational model of performance on the nine-dot and similar problems (MacGregor, Ormerod, & Chronicle, 2001), which is capable of more detailed predictions than previous cognitive approaches. The model is in the tradition of information-processing models of problem solving (Newell & Simon, 1972) and consists of three main components, comprising (1) scope of search , (2) move selection, and (3) move evaluation based on a criterion of progress towards the goal. The scope of search is governed by mental "lookahead"—the number of lines the person considers in advance. In the case of the nine-dot problem, this may range from a minimum of one to a maximum of four. Move selection is governed by the locally rational strategy of choosing the line(s) that intersect(s) the maximum number of dots, within the limits imposed by lookahead. Selecting moves that maximize the number of dots cancelled represents a form of "hill climbing" or "difference reduction" and is consistent with the view that humans make choices under conditions of bounded rationality (Chater & Oaksford, 1999; Gigerenzer & Goldstein, 1996; Simon, 1990). Move evaluation tests the move(s) suggested by this operator against a criterion of progress based on the ratio of the number of dots remaining to the number of lines remaining. The intuitive basis of this is the notion that the perceived progress supplied by a move will depend not just on the number of dots cancelled, but on their number relative to the number of dots and lines remaining. Thus, for example, a move that cancels one dot only may be unacceptable when many dots and few lines remain, but acceptable if few dots and many lines remain. The criterion we propose is that to be acceptable a move must cancel a minimum of the ratio of the remaining number of dots to lines. Criterion failure results when no acceptable move can be found and may lead to modification or abandonment of the primary operator.

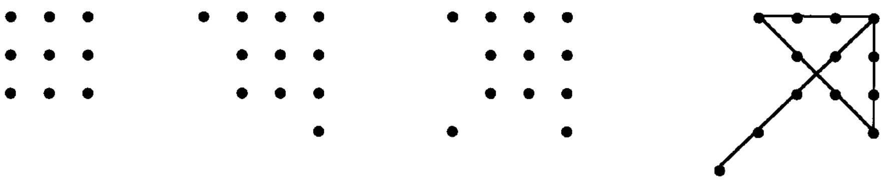
To exemplify these three components, consider a participant who is tackling the standard nine-dot problem shown in Figure 1a. Imagine that he or she has a lookahead value of two, and that he or she intends to start at the bottom left-hand corner dot. One of the most likely moves for this person to make is simply to draw round two sides of the dot array. This cancels five dots, and thus satisfies the criterion described earlier (as more than half of the dots are cancelled by using exactly half the number of lines available). Another participant operating at a lookahead value of three will, in all likelihood, draw round three sides of the dot array, thus cancelling more than three-quarters of the dots with three out of four lines. Both participants will, however, then find themselves unable to find a successful solution: They have been misled by the early success of their primary operator. Note that both these moves are apparent in data from human solvers: Furthermore, moves that do not conform to the model's predictions—such as going from the bottom left-hand dot, up one dot, then over to the middle dot—are only infrequently observed.

The model outlined predicts that anyone operating at less than four lookahead will fail to solve the nine-dot problem, at least on initial attempts. This is because, at lookahead values of one, two, and three, criterion failure does not occur early enough in the solution process to prompt a search for alternative, non-model, moves. The results for the nine-dot problem reported in our first experiment (MacGregor et al., 2001) supported these predictions, with 96% of classifiable first responses being consistent with the model. Model compliance dropped to 71% on the second attempt and to 53% on subsequent attempts, suggesting that repeated failures may lead to a search for other strategies. Nevertheless, the overall level of model compliance was high. The model also accurately predicted relative success rates for problems such as those shown in Figure 2 (MacGregor et al.).

In contrast to these positive predictions the present article is concerned with a set of negative predictions, where the model and its associated results suggest performance failures in the



Incomplete shapes



Complete shapes

Figure 2. Problems used in Experiment 3. The requirement for each problem is exactly as for the standard nine-dot problem shown in Figure 1a. The correct solution is shown only for the thirteen-dot problem; the solution is identical in all problems.

face of experimental manipulations that, on all other reasonable grounds, might be expected to greatly facilitate performance on the nine-dot problem. The article reports three experiments designed to test the importance of figural factors, in which visual cues to the solution of the nine-dot problem are provided in a manner designed to avoid the conceptual difficulties in earlier research identified previously. That is, the visual cues neither change the task requirement for participants, nor explicitly provide components of either the solution itself or of the "insight" required to find a successful solution (that is, drawing lines outside of the square boundary implied by the nine-dot array). Experiment 1 examines the effect of providing (1) a solution-relevant, Gestalt organization superimposed on the problem in the form of shading, and (2) the same superimposed shading with an additional written hint as to its relevance. Under both the traditional Gestalt account and our problem–solution mapping account (Ormerod et al., 1997), this manipulation should facilitate problem solution. Indeed, as the solution form is itself available as a percept, it should remove all aspects of problem difficulty. The computational model we recently proposed, however, would seem to predict little or no facilitation, because extending lines as implied by the shading is unnecessary to achieve maximum dot cancellation (except when the value of lookahead is at the maximum of four moves). The provision of the written hint in the second condition was suggested by the finding of Gick and Holyoak (1980) that hints are necessary to facilitate the use of solution-relevant information in some problem-solving situations. Experiment 2 is similar to Experiment 1, but employs a control condition in which a verbal hint to go outside the square shape of the problem was given, in order to directly compare the effects of verbal and perceptual hints. Experiment 3 uses a training-transfer paradigm to examine whether experience with the production of a directly transferable solution shape may successfully facilitate performance on the nine-dot problem.

EXPERIMENT 1

Method

Participants

A total of 87 people volunteered to participate in the experiment during a departmental open day. Of these, 4 did not follow instructions correctly, and their sheets were discarded. In the remaining 83, there were 38 men and 44 women, with ages ranging from 18 to 65 years (mean age: 45 years). One participant declined to give age and gender information.

Design

Two groups of participants were tested in two hint conditions: shading only and shading plus hint. A no-hint condition was not included in the design because of the robust and often-replicated floor effect with the nine-dot problem presented alone—instead we used previous results to estimate the expected frequency of solvers under the null hypothesis of no effect of hint.

Materials

Each participant was provided with a single sheet, on which were printed the problem instructions, problem and visual cue, written hint (in the second condition only), and questions about prior experience with the problem, plus spaces to record age and gender. The problem and visual cue are reproduced in Figure 1d. The problem instructions were as follows: "Your task is to draw four lines which, between them, go through all the nine dots. However, you must abide by the following constraints: the lines must be straight; the lines must be connected; once you have started drawing you must not lift your pen off the page; you must not retrace over a line you have already drawn." In the shading plus hint condition, the following was printed above the problem: "HINT: You may find it helpful to consider the shading under the nine dots when trying to find a solution."

Procedure

Both groups of participants were tested in a group setting in a university lecture theatre. Problem sheets for the two conditions were distributed in random order, and participants were asked to keep the sheet face down until told to turn over. The experimenter explained the general purpose of the experiment, and went through the problem instructions verbally, answering any queries that participants had. Participants were then told to commence the problem, and were given 3 min to complete it. Following the procedure participants were told of the purpose of the experiment and shown the solution.

Results

Of the 83 participants, 25 stated that they had seen the nine-dot problem and solution before: Data from these participants were excluded from further analysis. Of the remaining 31 in the shading only condition, 5 (16%) solved the problem, whereas 8 of 27 (30%) in the shading plus hint condition solved it. Analysis of the solution accuracy data showed no significant difference between the two conditions in the proportion of participants finding a correct solution, $\chi^2(1) = 1.51, p = .22$.

Further comparisons were made against expected frequencies of solution based on the results reported in previous published studies. Three of the published studies for which sample sizes were given reported zero solvers, based on samples of 15 (Burnham & Davis, 1969), and 15 and 12 (Weisberg & Alba, 1981, Experiments 1 and 1c, respectively). Two studies reported three solvers, based on samples of 32 and 33 (from Weisberg & Alba, Experiment 1a, and Lung & Dominowski, 1985, respectively). We took the average of these results as the population solution rate (5.6%) against which to test the goodness-of-fit of the present sample results, using the one-sample Kolmogorov–Smirnov test (the χ^2 test was not used because of small expected frequencies). For neither condition did the observed probability of solving significantly exceed the expected value ($D_{\mathrm{MAX}} = .11, p > .20$ for the shading only condition, and $D_{\mathrm{MAX}} = .24, p > .05$ for the shading plus hint).

Failed solution attempts were examined for information regarding the problem-solving process. Of the 26 failed attempts in the shading only condition, 10 (38%) were blank. Similarly, of the 19 in the shading plus hint condition, 9 (47%) were blank. The remainder were classified in two ways: First, in terms of whether or not the solution attempt was consistent with the computational model, and second, in terms of whether or not any attempt at a line extension beyond the square shape of the nine dots had been made. Here and in the following experiments, we report percentages based first on classifiable attempts only, then on all responses (including blanks and unclassifiable attempts) in the format aa%/bb%. Statistical tests are based on the former. In the shading only condition, eight attempts (50%/26%) were consistent with the model, and in the shading plus hint condition, five attempts (50%/26%) were consistent. Of all the non-model-consistent attempts in both conditions, five extended a line beyond the square: three (19%/12%) in the shading only condition and two (20%/11%) in shading plus hint. The difference was not significant, $p = .57$, Fisher exact test.

Discussion

The results of Experiment 1 provided little or no evidence that, when given a visual cue as to the shape of the solution, participants are more likely to solve the nine-dot problem than with no visual cue present. The shaded solution shape seems, on the face of it, to provide an unambiguous pointer to the solution, and it is perhaps surprising that so few participants capitalized upon it. With an explicit written hint to use the cue to construct the solution, success rates improved slightly, but not to a degree that was significantly greater than the proportion in the shading only condition or, for that matter, the proportion expected with no hint at all, based on previous research. At an intuitive level, it remains surprising that only 30% of participants found a solution in this case even when (1) the shape of the solution was perceptually available (2) the written hint clearly indicated the relevance of this alternative shape information. These data tend, on the face of it, to argue against our previous suggestion (Ormerod et al., 1997) that problem/solution shape incongruence underlies the difficulty of the nine-dot problem. They are, however, consistent with our more recent computational approach (MacGregor et al., 2001).

A more detailed analysis of failed solution attempts further supported the conclusion that there were little or no differences between the two conditions. In many cases participants simply failed to produce any response. However, of those who did, 50% in each condition generated solution attempts that were consistent with the computational model, whereas relatively few extended lines beyond the boundary of the nine dots. We comment further on the observed level of model compliance in the General Discussion.

One disadvantage with the design of Experiment 1 is that no direct comparison was possible between the visual cues we used and the verbal "exhausted square" hint used by Weisberg and Alba (1981). Although the facilitation provided by the shaded solution shape was surprisingly small, it remains possible that some aspect of the particular stimuli or procedure prevented participants from successfully finding a solution. In order to address this potential concern, Experiment 2 was conducted.

EXPERIMENT 2

Method

Participants

A total of 149 people volunteered to participate in the experiment during several departmental open days. There were 63 men and 84 women, with ages ranging from 16 to 62 years (mean age: 42 years). Two participants declined to give gender information, and three declined to give age information.

Design

Three groups of participants were tested in three hint conditions: (1) written hint to go outside the square shape of the problem, henceforth the exhausted square condition; (2) shading plus a written hint to use the shading as an aid (identical to Condition B in Experiment 1, henceforth the shading plus hint condition); (3) shading plus a written hint to use the shading, plus a written hint to go outside the square shape of the problem (the shading plus both hints condition).

Procedure

Problem sheets were very similar to those used in Experiment 1. In the exhausted square condition, the shaded solution shape was omitted, and the following written hint presented above the problem: "HINT: Solution attempts whose lines remain entirely within the square shaped pattern made by the nine dots *cannot* succeed." The shading plus hint condition had the shaded solution shape and the following written hint: "HINT: You may find it helpful to consider the shading under the nine dots when trying to find a solution." The shading plus both hints condition had shading plus both of the written hints. In other respects the procedure was identical to that used in Experiment 1.

Results

Of the 149 participants, 39 stated that they had seen the nine-dot problem and solution before: Data from these participants were excluded from further analysis. Of the remaining 110 there were 44, 27, and 39 in the exhausted square, shading plus hint, and shading plus both hints conditions. The number (percentages) solving in these conditions were 7 (16%), 7 (26%), and 13 (33%), respectively. There was no overall significant difference among conditions, $\chi^2(2) = 3.43, p = .18$.

As in Experiment 1, failed solution attempts were examined in detail. Of the 83 failed attempts, 41 (49%) were blank. The remaining failed attempts were classified for model consistency and, overall, 60%/30% were consistent with the model. Broken down by condition, the rates were 45%/24%, 75%/30%, and 71%/38% for the exhausted square, shading plus hint, and shading plus both hints conditions, respectively. The conditions did not differ significantly in this respect, $\chi^2(2) = 2.40, p = .30$. There were six cases of failed solution attempts that extended a line beyond the square, four (20%/11%) in the exhausted square condition, zero in the shading plus hint condition, and two (14%/8%) in shading plus both hints condition. The difference between conditions was not significant, $\chi^2(2) = 1.87, p = .39$.

Discussion

The exhausted square and shading plus hint conditions gave rise to a comparable pattern of performance; it is further notable that the exhausted square condition gave rise to similar levels of performance when compared with the equivalent condition from Weisberg and Alba (1981), where 4 of 15 participants solved the problem. In addition, the results replicated those of Experiment 1 for the shading plus hint condition, with 26% solving compared with 30% in Experiment 1. As with Experiment 1, a potentially surprising feature of the data was that solution rates were not considerably higher, particularly in the shading plus both hints condition, where it could be argued that participants have explicit hints for cognitive strategy (to go outside the square shape), implicit hints to overcome fixation on the Gestalt of the problem (the shaded solution shape), *and* an explicit instruction that the anti-fixation hint might be useful, and as such ought to be solving the problem with near-100% success. However, the clear prediction of our computational model is that in all conditions of Experiment 2, neither shading nor hints will influence performance, as participants are dominated by the initially successful operator of maximum dot cancellation. Consequently, for example, drawing the first three lines of a solution around three sides of the square formed by the nine dots cancels seven dots, the maximum possible. There is therefore no motivation to consider using the information provided by the hints to seek moves that extend beyond the dot array.

It is, however, possible to suggest that a remaining component of the problem's difficulty lies in the execution of the drawn solution. That is, to achieve insight into the solution of the nine-dot problem requires not only the provision of solution hints, but also the experience of the procedures involved in their application. A third experiment was conducted to investigate this possibility.

EXPERIMENT 3

In an attempt to provide participants with experience of producing a solution exactly akin to the solution required for the nine-dot problem, Experiment 3 trained participants on three problem variants with an identical solution shape, which we know from previous research can be reliably solved by naïve participants (MacGregor et al., 2001). These problems should therefore provide suitable production experience. Thus the experiment employed a training and transfer design. Participants first received three problems in which they were trained to produce the solution shape of the nine-dot problem using eleven-, twelve, and thirteen-dot problem variants. They received the standard nine-dot problem as the transfer task.

The stimuli used are shown in Figure 2. In the lower panel, the eleven-dot problem consisted of the original nine-dot problem plus two additional dots, one located one unit to the left of the upper-left-hand dot and a second dot one unit below the bottom-right-hand dot of the original nine-dot pattern. The twelve-dot problem was next generated from this eleven-dot version by adding one more dot one diagonal unit from its bottom-left-hand dot. Finally (for the lower panel) the thirteen-dot problem was obtained by adding one more dot one diagonal unit from the bottom left of the twelve-dot pattern. Each of these problems has dots located at the inflection points at the head of the familiar arrow-shaped solution of the nine-dot problem, and they are henceforth termed "head complete" problems. The training stimuli shown in the upper panel of Figure 2 were next obtained by removing dots from the "head" of this arrow (henceforth termed the "head incomplete" problems). That is, the twelve-dot problem in the upper panel was produced by deleting the bottom-right-hand dot from the thirteen-dot problem. Similarly, the eleven-dot problem was obtained by deleting the top-left-hand dot from this twelve-dot version. (Note that the thirteen-dot problem was identical in both conditions.)

Our previous work with problem variants such as those shown in Figure 2 (MacGregor et al., 2001) suggested that the head complete and head incomplete conditions would provide an effective manipulation of the number of times a correct solution shape would be produced. In the head complete conditions shown in the lower panel in Figure 2, participants were expected to solve reliably by drawing the arrow-shaped solution on to the dots provided. In contrast, for the eleven- and twelve-dot head incomplete problems, participants were not expected to draw the arrow-shaped solution. If, as suggested earlier, a residual source of difficulty in the nine-dot problem is that participants require experience of executing the solution in order to solve it, then it would be predicted that the head complete conditions would lead to higher levels of transfer than would the head incomplete conditions. A further factor of order was also included, with training problems either being built up from eleven to thirteen dots across problems or reduced. Order was included to test whether transfer would be more likely when training problems became successively better approximations, from the point of view of perceptual similarity, to the nine-dot transfer problem. It should be noted that the number of dots forming the *tail* of the arrow shape of these problems does not influence performance despite seemingly providing a cue as to a starting point (MacGregor et al., 2001, Experiment 2).

In contrast to the preceding predictions, our computational model (MacGregor et al., 2001) predicts little or no transfer to the nine-dot problem. For the head complete training problems participants will, in fact, receive reinforcement if they perform according to the maximum dot-cancellation operator of the model, as this will generally be accompanied by successful completion of those problems. Thus they should continue to be dominated by this operator on the nine-dot problem, which more or less guarantees that a correct solution will not be found (except in those cases where criterion failure occurs, which in turn requires a lookahead of four lines). For the head incomplete training problems, the model similarly predicts little or no transfer to the nine-dot problem. Participants will remain consistent to the model and generally attempt to maximize the number of dots cancelled per move: This will result in only low levels of correct solutions to the eleven- and twelve-dot training problems.

Method

Participants

The participants were 188 undergraduate psychology student volunteers. As the experiment was carried out on a group basis with limited time available, identifiers and age information were not collected.

Design

The factors manipulated were: (1) problem, thirteen-dot, twelve-dot, eleven-dot, or nine-dot; (2) shape, either head complete or head incomplete; and (3) order of training trials, either 13–12–11 or 11–12–13. Factor (1) was assigned within participants and Factors (2) and (3) between groups.

Materials

Each participant received a printed problem booklet. The first page gave general instructions relevant to all four problems; the instructions were identical to those used in Experiment 1. The following four pages contained the four problems in the appropriate order for the experimental condition. The two shape conditions for the eleven-, twelve-, and thirteen-dot problems are shown in Figure 2. A final page asked some debriefing questions about participants' prior knowledge of the nine-dot problem.

Procedure

Testing of participants was undertaken as part of undergraduate lecture classes. Problem booklets were distributed to participants, and the instructions were outlined verbally by the experimenter, any queries being answered at this time. Participants were given 3 min for each problem, and were asked to turn the page to the next problem by the experimenter at the correct time intervals. At the end of the testing period, the purpose of the experiment and the correct solution were explained to the group of participants.

Results

A total of 70 participants indicated that they had seen the nine-dot problem on a previous occasion, and their solutions were therefore excluded from further analysis. Each solution to each problem from the remaining 118 participants was scored 1 for a correct solution and 0 for incorrect.

The main results are illustrated in Table 1, which shows the percentage of participants producing a correct solution for the 16 conditions. Overall, the percentages of participants solving the thirteen-, twelve-, eleven-, and nine-dot problems were 71%, 43%, 29%, and 5%, respectively. This overall pattern of results was consistent with our previous findings (MacGregor et al., 2001).

As a manipulation check, the total numbers of times that participants produced solutions of the correct shape (the standard arrow shape shown in Figure 2) were submitted to a between-participants analysis of variance, with pattern and order as the factors. As expected, there was a highly significant effect of pattern, with head complete problems giving rise to many more correctly shaped solutions (mean = 2.18) than head incomplete problems (mean = 0.88); $F(1, 114) = 49.3$; $p < .001$, $MSE = 0.95$. Neither the main effect of order nor the interaction between pattern and order was significant.

Performance on the nine-dot test trial was also analysed using a between-participants analysis of variance, with pattern and order of training as the factors. The results indicated no significant effects (all *p*s > .05), and performances were low across all four conditions (Table 1).

Failed responses to the nine-dot problem were classified in terms of whether or not they were consistent with the computational model. There were a total of 88 classifiable responses (unclassifiable responses were either blank or ambiguous), of which 61 (69%/54%) were model consistent. For the head complete shapes, the number (percentages) of model consistent responses were 8 (80%/47%) and 21 (72%/58%), for the 13–11 and 11–13 orders of training, respectively. The comparable results for the head incomplete conditions were 13 (52%/42%) and 19 (79%/68%). The results were analysed using a 2 × 2 between-groups analysis of variance, which indicated no significant effect of shape, order of training, or their interaction (all *p*s > .05).

The results were further analysed in terms of the number of failed solutions to the nine-dot problem that exhibited a line extension beyond the square outline of the dots. Of the 93 completed responses, 14 showed line extensions. For the head complete shapes, the frequencies (percentages) of line extensions were one (8%/6%) and two (7%/6%), for the 13–11 and

TABLE 1
Percentage of solvers as a function of the three experimental factors in Experiment 3

		Problem				
Shape	*Order*	*13*	*12*	*11*	*9*	*N*
Complete	13 to 11	78	89	78	6	18
Incomplete	13 to 11	69	9	9	3	32
Complete	11 to 13	81	72	39	0	36
Incomplete	11 to 13	59	19	9	13	32

11–13 orders of training, respectively. The comparable results for the head incomplete conditions were eight (31%/25%) and three (12%/9%). A 2 × 2 between-groups analysis of variance indicated no significant effects (all *p*s > .05).

Discussion

Solution rates to the nine-dot problem were low across all four combinations of training factors and were not significantly different from each other. The training manipulation gave rise to significantly more experience with correct solution shape production in the head complete conditions, as expected. The lack of transfer of the shape of the solution from the training problems to the test problem in the head complete conditions argues strongly that simple experience with solution shape does not make the nine-dot problem easier to solve. This is so despite (1) the identical shape of the solution in all cases; (2) the fact that the nine-dot problem array is embedded within each of the problem variants, and (3) the fact that in the 13–11 condition, the order of training trials let the participants experience problems that were progressively more and more similar to the transfer problem. This failure to transfer experience with the solution shape to the nine-dot problem is consistent, however, with the prediction of our computational model. As criterion failure will be rare (limited to people operating at a lookahead of four lines) there is no impetus to entertain "non-model" moves, such as line extensions, thereby guaranteeing failure.

A more detailed test of the model's predictions was again made by examining the percentage of model-consistent solutions to the nine-dot problem. Overall, 69%/54% of responses were consistent with the model, and although there were more in the head complete than head incomplete conditions, the difference was not statistically significant. A possible criticism of the experimental method used in Experiment 3 is that participants may not have been sufficiently aware of the spatial relationship between the training and test problems. Indeed, the results we report in Ormerod et al. (1997), where participants failed to solve an eleven-dot problem in which the standard nine-dot problem was made distinct, tend to support this view. However, this experiment was a one-trial study, in which participants did not benefit from any transfer or instructional manipulation. Therefore, in order to examine this point in the context of a training-transfer paradigm, a further 25 participants were presented with the training and test problems shown in Figure 3. The first training problem was the standard thirteen-dot problem used in Experiment 3, and the second training problem had the relevant nine dots

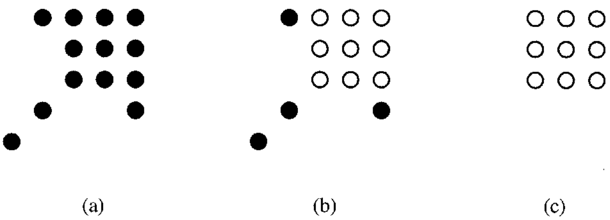

(a) (b) (c)

Figure 3. Supplementary problems for Experiment 3. Panel (a) shows the first training problem, panel (b) the second training problem, and panel (c) the standard nine-dot transfer problem. The solution is in all cases identical to that shown in Figure 2.

distinguished by being unfilled circles. The test problem consisted only of the nine unfilled circles in the normal nine-dot pattern. Solution rates were 88, 84, and 12% respectively. This pattern of data demonstrates that the failure to transfer in Experiment 3 was not due to confusion over how the test problem was embedded in the training problem: When this relation was made transparently clear, the failure to transfer remained.

GENERAL DISCUSSION

In this paper we have argued that previous attempts to study the operation of hints in the nine-dot problem have been problematic because they qualitatively changed the problem-solver's task. In addition, they have generally focused on strategic types of hint presented at a cognitive level. In the experiments reported here, attempts were made to provide visual and perceptual–procedural cues to a solution, without varying the task, and without providing elements of the solution itself. It appears from all three experiments, however, that these cues to the solution do not provide a great deal of facilitation. In general, therefore, it might be concluded that experiencing the shape of the solution is not in itself sufficient to prompt application of that shape to the nine-dot problem.

These findings are of considerable interest from a theoretical point of view. The nine-dot problem has frequently been regarded in the literature as a problem that requires *insight* for a solution. Ohlsson (1992, p. 4) defines insight as occurring ". . . in the context of an impasse which is unmerited in the sense that the thinker is, in fact, competent to solve the problem". In the nine-dot case, adult solvers are undoubtedly competent to solve, but they encounter an impasse, which, again in Ohlsson's terms, has usually been regarded as being breakable only by the process of constraint relaxation. That is to say, the inferred constraint on the part of participants to remain within the square shape implied by the array is relaxed, so that lines may extend beyond the figural boundary, and a solution may be found. In the experiments reported here, we would argue that the visual cues provided guide the participants strongly towards constraint relaxation—and yet the impasse is broken only in a small minority of cases, judging by the solution rates. We therefore suggest that it may be worthwhile to consider the process of insight in this problem in a slightly different light.

The results of these experiments do not appear to be consistent with either the traditional Gestalt account or our problem–solution mapping account (Ormerod et al., 1997). Furthermore, the general failure of quite directive hints to encourage participants to use non-dot points in constructing a solution is at odds with the theoretical position of Weisberg and Alba (1981) and Lung and Dominowski (1985). Rather, the findings reported here appear to be more consistent with our recent computational modelling approach (MacGregor et al., 2001), which proposes an operator that seeks to "cancel" as many dots in the problem as possible with each solution line. In the model, a solution to the nine-dot problem can only be achieved if the average dot cancellation by each line equals 2.25—this criterion, in turn, can only be achieved if there is sufficient lookahead to find the appropriate solution. According to the model, provision of visual cues to the problem's solution will not affect the application of the cancellation criterion and so should not influence performance on the standard nine-dot problem, where the criterion is never violated. The model also has advantages over previous cognitive accounts in being able to explain both success and failure in problems that do not require non-dot point inclusion.

The overall rates of model-compliant responses to the nine-dot problem ranged from 45%/24% to 80%/47% in the present experimental conditions. These figures are substantially lower than the figure of 96% found previously for a first attempt, and are more in accordance with the rates we have reported of 71% for a second attempt and 53% for subsequent attempts (MacGregor et al., 2001). A possible reason for this difference is that the latter study gave participants 1 min for each of ten attempts, whereas the present experiments allowed 3 min for one attempt. It may be that in the present case some participants' nominal first attempt may effectively have been a second or third attempt because they used the additional time to try to discard moves mentally.

To summarize, we would claim that it is reasonable to regard the processes of criterion failure and lookahead embodied in a hill-climbing model of nine-dot problem solving as belonging to the route to insight that Ohlsson (1992) terms "re-encoding". In other words, a fundamental reconceptualization of how the problem may be tackled is required: Constraint relaxation in and of itself is not sufficient to promote successful solutions. Furthermore, it is clear that visual hints designed to aid constraint relaxation, such as those examined in the foregoing experiments, are unlikely to impact upon the development of operator lookahead, as lookahead is usually thought of as being constrained more by working memory or other capacity limitations than by perceptual factors. In general, therefore, the findings reported here make it very clear that the answer to a problem may be capable of eluding human solvers even when it is literally staring them in the face—and, furthermore, when they are explicitly told that it is staring them in the face.

REFERENCES

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

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*, 57–65.

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

Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Groome, D. (1999). *An introduction to cognitive psychology: Processes and disorders.* Hove, UK: Psychology Press.

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*, *111*, 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,* 176–201.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

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

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

Ormerod, T.C., Chronicle, E.P., & MacGregor, J.N. (1997, August). *Facilitation in variants of the nine-dot problem: Perceptual or cognitive mediation?* Poster presented at the 18th Annual Conference of the Cognitive Science Society, Stanford, CA.

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

Simon, H.A. (1990). Invariants of human behaviour. *Annual Review of Psychology*, *41*, 1–19.

Wade, C., & Tavris, C. (1990). *Psychology* (2nd ed.). New York: Harper and Row.

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.

Original manuscript received 8 October 1999
Accepted revision received 25 July 2000