Information-processing and insight: A process model of performance on the nine-dot and related problems.

James N. MacGregor

University of Victoria, Canada.

jmacgreg@hsd.uvic.ca

Thomas C. Ormerod, Edward P. Chronicle

Lancaster University, UK.

{T.Ormerod, E.Chronicle}@lancaster.ac.uk

Journal of Experimental Psychology: Learning, Memory & Cognition. Vol 27, 176-201.

Abstract

The nine-dot problem is widely regarded as a difficult *insight* problem. Current explanations of its difficulty do not provide process models, and it has been argued that insight problem-solving cannot, in principle, be specified within an implementable model. Here, we present such a model based on maximization and progress-monitoring heuristics with lookahead. In Experiments 1 and 2, the model predicted performance over trials for the nine-dot and related problems. Experiment 3 supported an extension of the model that accounts for insightful moves. Experiments 4 and 5 provided a critical test, with the standard problem, of model predictions versus those of previous accounts. On the basis of these findings, we claim that insight problem-solving can be modeled within a means-ends analysis framework. Maximization and progress-monitoring heuristics are the source of problem difficulty, but also create the conditions necessary for insightful moves to be sought. Furthermore, they promote the discovery and retention of promising states that meet the progress-monitoring criterion and attenuate the problem space.

Number of characters and spaces: 959

The nine-dot problem is one of the best-known stimuli from the psychology of problem-solving. It is shown in Figure 1 with its solution. The goal is to connect all nine dots by drawing four straight lines, without either retracing a line or lifting pencil from paper. The problem is simple to state and to represent, but it is extremely difficult to solve. The solution requires extending lines to unmarked locations beyond the implied boundary of the dots, and is widely regarded to involve *insight* (Isaak & Just, 1995; Ohlsson 1990), that is, the recognition or restructuring of a key feature of a problem that allows a solution to be found. A number of competing explanations for the problem's difficulty have been proposed (e.g., Sheerer, 1963; Weisberg and Alba, 1981a & b; Lung and Dominowski, 1985). All of these account for the failure to achieve insight into the solution as resulting from the inappropriate imposition of constraints beyond those given in the problem statement. Despite these explanations, controversy remains as to the precise source of difficulty in the nine-dot problem. More generally, the relationships, in terms of the cognitive processes or heuristics involved, between the solving of insight problems and of non-insight problems are considerably underspecified.

This paper builds upon existing explanations of nine-dot performance by developing and empirically testing a model that places the nine-dot problem squarely within a class of problems that can be understood in terms of the application of a means-ends-analysis heuristic. The model makes four key theoretical contributions: First, it generalizes from the problem-specific factors identified in previous accounts of the nine-dot problem, demonstrating that mechanisms which account for performance on non-insight transformation problems (e.g., Newell & Simon, 1972; Thomas, 1974; Simon & Reed, 1976) also underpin nine-dot performance. Second, it provides a detailed information-processing model of performance on a classic insight problem, which extends existing qualitative accounts of problem performance. Third, it offers a rigorous conception of "impasse" in insight problem-solving, as well as explaining how the conditions necessary for insight arise, and why insight is so rare in the standard nine-dot problem. Fourth, it introduces the concept of 'promising states', that is, novel moves that (i) are made once the conditions for insight arise and (ii) lead to a realisation that these types of move are of potential value, thus allowing a reconceptualisation of the problem space.

The traditional explanation for the nine dot problem's difficulty is a Gestalt one, that people fixate upon the square shape implied by the dot array and consequently fail to consider possible solutions that fall outside its boundary (Scheerer, 1963). Several investigations attest to the problem’s difficulty, reporting solution rates of 0% (Burnham and Davis, 1969; Weisberg and Alba, 1981a, Experiment 1). Both groups of researchers also investigated the effect of explicit instructions to violate the implied array boundary: rather little facilitation was observed (thus raising, though not answering, the second question outlined above). These results have been interpreted as falsifying the Gestalt account, since even when instructed to seek solutions outside the implied square shape, few solved. The validity of this conclusion has since been questioned (Dominowski, 1981; Ellen, 1982; Lung and Dominowski, 1985; Ohlsson, 1990). Nonetheless, a number of alternatives to the Gestalt account have been proposed that emphasize cognitive rather than perceptual factors. Weisberg & Alba (1981a) proposed a hypothesis-testing explanation in which they assumed that people would initially sample hypotheses from a domain based on previous experience with “connect the dot” types of problems. This domain, while it does not contain the correct solution, may be sufficiently large that participants do not exhaust it, and consequently do not solve the problem. Lung and Dominowski (1985) found difficulties with this framework, pointing out that the domain of incorrect solutions indicated by Weisberg and Alba’s experimental results was in fact quite limited. They argued it is the imposition of an inappropriate constraint, that all lines must begin and end on dots, that makes the nine-dot problem difficult. Subsequently, explanations of this general type, that base the difficulty of the problem on the imposition of inappropriate or superfluous operator constraints, have been proposed by a number of authors (Isaak & Just, 1995; Ohlsson, 1990).

These post-Gestalt theoretical attempts share one characteristic with the Gestalt accounts, that they are essentially qualitative. Their main prediction is that the nine-dot problem will be difficult, but without specifying how difficult. If they suggest manipulations that will reduce difficulty, they are unable to predict by how little or how much. In this respect, they differ markedly from theories that appeared during the same general period which were able to predict performance with other “move” problems, such as the Waterjug, Missionaries and Cannibals and Towers of Hanoi problems. For these problems, detailed process models based on the general information processing approach described by Newell and Simon (1972) have been proposed (e.g. Atwood & Polson, 1976; Jeffries, Polson, Razran & Atwood, 1977; Karat, 1982). No similar model exists of performance on the nine-dot problem. Indeed, it has been questioned whether such a model could be developed (Weisberg & Alba, 1981a; Wertheimer, 1985).

The apparent difficulty of modelling the nine-dot problem stems from the nature of its goal state, which is defined in abstract terms only (Weisberg and Alba,1981a). This contrasts with well-defined problems such as the Towers of Hanoi, where the goal state is necessarily concrete and visualizable ("move the disks from start to destination pegs"), At first sight, it is unclear how traditional information-processing heuristics such as means-ends analysis and goal reduction can be applied to tasks like the nine-dot problems. If solvers do not have a concrete and visualizable “end” against which to monitor their progress, how can they apply operators that will reduce the “distance” between the current state and the goal state? We propose that they nevertheless do so, by employing two related general principles. The first is to seek and apply a “locally rational” operator that reduces the distance between the current state and some local sub-goal state. The second is to apply a criterion against which to monitor progress through the problem. If at each stage the locally-rational operator meets this criterion, then it will be applied. If not, the problem space expands and a search for alternative operators takes place. The criterion also provides a bench mark against which to measure alternative operators. A new operator that meets or surpasses the current criterion may be judged to be “promising” and repeated. One that does not may be discarded. The importance of these two general principles lies in their ability to account for both failure and success in problems with abstractly-defined solutions. Adherence to an inappropriate locally rational operator creates a source of difficulty in such problems. Criterion failure and the subsequent search for an alternative operator create conditions under which a solution may be achieved.

Model description - Stage 1

In Stage 1 of our model, we suggest that a likely locally rational operator for people to select when faced with the nine-dot problem is to choose lines that intersect the maximum number of dots. This proposal is not new. Burnham and Davis (1969), for example, suggested that people might tend to select lines that join most dots. Lung and Dominowski (1985) refer to "optimal moves" as those connecting the greatest number of dots, and outlined a recursive three-stage information-processing approach that involved selection of a dot to start a move, evaluation of moves from that dot and selection and execution of a move. They observed that participants tended to select lines that connected the maximum number of available dots. The results of their experiment indicated that 95% of all first lines intersected three dots, while 83% of second lines intersected two, the maximum possible. Selecting moves that maximize the number of dots cancelled represents a form of "hillclimbing" 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), and is implemented in Stage 1 through the definitions and assumptions described below. Full details for each of the problems used in the experiments reported here are given in the Appendix.

Definitions

1. Adjacent dots are dots that are 1 unit or 2 units apart (i.e. one lateral or diagonal unit).

2. A line consists of drawing a *straight* line intersecting a series of two or more adjacent dots (Note: this is functionally equivalent to the Gestalt view of the nine-dot problem that people do not initially consider moves that lie beyond the perimeter of the dots).

3. A move consists of a sequence of one or more successive lines (such that a move consisting of N lines will incorporate N-1 changes of line direction).

4. The value of a line is given by the number of previously unconnected dots it intersects.

5. The value of a move is the sum of the values of its component lines.

Assumptions

1. At each step the move with the highest value is selected from the array of legal possibilities. If there is more than one with the maximum value, selection between these is made randomly.
2. Previously intersected dots have no value but may continue to act as place markers. That is, lines may be extended from and to previously intersected dots, but these will not add to the value of the move.
3. The evaluation of possible moves may consider moves of one or more lines up to a maximum of the number of lines remaining depending on the adopted strategy or memory capacity of the individual ("lookahead").
4. The lookahead value remains constant across a solution attempt.

These model definitions and assumptions constitute what we propose are the constraints under which human solvers operate.

Example

To illustrate, consider the eleven-dot problem used by Burnham and Davis (1969), but as shown in Figure 2. In the lower panel of the figure, dots have been replaced by numbers to facilitate exposition. Consider first an individual with a lookahead value of one line. (For simplicity, Assumption 4 considers the lookahead value to remain constant within one solution attempt.) The best available first moves intersect four points (Assumption 1), and there are six such moves (1 to 10, 1 to 11,10 to 11, and the reverse of these). Only two can lead to a correct solution (1 to 10, 1 to 11). Assuming these choices to be equiprobable (Assumption 1), then the likelihood of a correct first move is 2 out of 6. Following the fate of an initial move from 1 to 10, the second move will be from 10 to 11, the single best move available and intersecting three adjacent points (Assumption 2), making the likelihood of a correct second move equal to 1. From this point, the best available moves intersect 2 points, either from 11 to 1, which lies on a correct solution path, or from 11 to 4, which will not lead to a correct solution. With a lookahead limited to the current move, the model assumes that these moves are equally likely, and consequently the conditional probability of a correct third move, given that the first two moves were correct, is 0.5. If a correct third move is chosen, the best available fourth move, from point 1 to point 9 solves the problem. A similar ratio of correct and wrong moves would follow if the operator had selected the other correct first move, from point 1 to point 11. Consequently, the probability of a move being a component of a correct solution is 0.33, 1, 0.5 and 1 for moves one through four, respectively. The overall probability of being correct (on the first attempt) is their product, 0.17.

Next, consider an operator that evaluates moves consisting of two lines at a time (a lookahead of two). In this case the best first moves have a value of seven, and again there are six such moves. Once again, two of them lie on a correct solution path. These are the lines 1-10-11 and 1-11-10. The wrong first moves are 10-1-11, 11-1-10, 10-11-1 and 11-10-1. Following the fate of the first of the correct initial moves, the best second move connects points on the path 11 through 1 to 9 and completes the solution. Similarly, the other correct first move, 1-11-10 leads inevitably (according to the model) to a correct second and final move. The overall probability of solving with a lookahead of two is therefore 0.33.

With a lookahead value of three lines, solving the problem will take two moves. The best first move consists of three lines intersecting 9 points, and there are 14 such moves. Four of these (1-10-11-1, 1-11-10-1, 9-1-10-11 and 9-1-11-10) are on a correct path, while the remainder are not. With equiprobable moves, the likelihood of a correct first move is 0.29. Note that some of the wrong first moves have the same characteristic of stopping short as those that occurred with the one lookahead operator. For instance, the correct first move, 1-11-10-1, has a closely related wrong move, 1-11-10-2. Following the fate of correct first moves, the best second move consists of one line intersecting two points, and there are 6 such moves. Having made either of the first moves 1-11-10-1 or 1-10-11-1, the only maximal move available is to draw the line connecting point 1 to 9, thereby solving the problem. The correct first move 9-1-11-10 leaves the operator with a choice of two possible second moves, 9-1-11-10-1 and 9-1-11-10-2, both of which are correct. (The path 9-1-11-10-2 is considered correct since it intersects all dots with four straight lines. The problem statement does not require that the solution has to form a closed figure.) The same is true for the correct first move 9-1-10-11. Therefore, given a correct first move, there are six possible second moves, all of them correct. The conditional probability of being correct on move two given a correct first move is therefore one, and their combined probability, 0.29.

Finally, with a lookahead of four lines, the maximal value of the first (and only) move is 11. There are six such moves, all of which result in correct solution (9-1-10-11-1, 9-1-10-11-4, 9-1-11-10-1, 9-1-11-10-4, 1-10-11-1-9 and 1-11-10-1-9). The probability of correct solution with a lookahead of four lines is therefore 1.00. In summary, the model predicts that the solution probabilities on the first attempt will be 0.17, 0.33, 0.29 and 1.00 for lookahead capacities of one, two, three and four lines, respectively.

Often, participants in experiments are provided with more than one attempt to solve, and there are a number of ways the model could be adapted to account for performance following an initial failed attempt. For example, moves could be selected randomly from the available optimal moves. Alternatively, selection may not be equiprobable across the set of optimal moves because, for instance, some optimal moves may be perceptually more salient than others. Moves, whether randomly-selected or not, may be sampled with or without replacement across successive trials, or in some intermediate way, where failing move on one trial may have their probability of selection reduced on subsequent trials. As a first approximation we have made the simplifying assumption that participants select randomly from the available optimal moves, with replacement. Under this assumption a sequence of attempts over trials will represent a series of Bernoulli trials where, if p is the probability of success on any attempt, then the expected number of trials to solve will be 1/p and the proportion of participants expected to solve within n attempts will be $1-(1-p)^n$. We recognize that this assumption is likely to be an over-simplification, but in the absence of other information it provides a reasonable baseline position.

No attempt has been made to elaborate on the psychological processes underlying the recognition of the longest "line". In the experiments described later, the maximum number of dots in any line was six, the minimum, three. We assume that people can discriminate which "line" of dots is longest in arrays of these kinds with relatively high levels of accuracy. Furthermore, we assume that the number of dots in a line is given to them through a process of subitizing. That is, we do not propose that they actually count dots.

The assumptions governing lookahead are, for the time being, simple. In particular, the proportions of participants using different values of lookahead are impossible to specify in advance. In Experiments 1 and 2, we have performed separate analyses at each value of lookahead, to quantify the success of Stage 1 of the model. The data from these two experiments were then used to estimate the proportions of different lookaheads in order to generate more detailed predictions from Stage 2 of the model, for subsequent experiments.

Stage 2

Because of the way in which it constrains the problem space, Stage 1 of the model leads inevitably to failure on the standard nine-dot problem, giving rise to an apparent impasse after a number of problem attempts. To allow for the possibility of success, the model is extended to a second stage where its constraints may be relaxed, permitting a wider (though not infinite) search space. Following the search theory of insight proposed by Kaplan and Simon (1990) we recognize two key issues in this extension – first, what triggers the shift from the initial problem space to a better or alternative one (i.e., what are the conditions necessary for insight to arise) and, second, what influences the formulation of better moves (i.e., how is an insightful solution implemented)?

In answer to the first question, Kaplan and Simon (1990) proposed that if "no operators seem to yield progress, one must search for a new problem space to explore…" (p.377). Note that this requires some judgment or evaluation of the progress gained by the current operators. This raises the question of how, in the nine-dot problem, a participant might judge the degree of progress. There are several possibilities. One is through the experience of failure, which provides a major retrospective cue as to the adequacy of current operators. In experiments that involve repeated trials, failures may provide the trigger for a change in the problem space. With problems like the matchstick algebra of Knoblich,Ohlsson, Haiber & Rhenius (1999), the experience of failure itself is the likely trigger, since the problems have single-step solutions. However, solution attempts in the nine-dot problem require multiple steps. As a consequence, concurrent progress monitoring, or the anticipation rather than experience of failure, has value. In part, its value is in allowing participants to modify or abandon solution attempts prior to failure itself. More importantly, it enables participants to gain information concerning the potential utility of a solution approach even if the current attempt fails. Participants' solution attempts occasionally generate promising states which result from move combinations that, although not immediately leading to solution, contain potentially useful elements (e.g., line extensions to non-dot points). If experience of failure were the only form of progress monitoring available to participants, then they would be unable to capitalize in future solution attempts upon promising states. Thus, concurrent progress monitoring is essential, not only to the operation of the maximization heuristic, but also for the recognition and maintenance across trials of potentially insightful moves. The retention of promising states that are executed succesfully over a number of trials is consistent with Ohlsson's (1990) suggestion that achieving insight may not be a one-step "aha" experience, but the result of and incremental cognitive process.

There is evidence that people subjectively monitor their progress during, as well as after, attempted problem solving (VanLehn, 1989). There are two ways in which this kind of prospective evaluation could be accommodated by our model. One is from the application of sufficient mental “lookahead” to anticipate whether a series of moves will eventually lead to success or failure. This is analogous to the experience of success or failure, except that it is provided by a "mental simulation". This would provide a definite answer but is also the most costly evaluation strategy in terms of cognitive load. A second is by evaluating prospective moves against some kind of internal criterion, providing a probabilistic estimate. For example, some of the protocols cited by VanLehn (1989) indicated that problem-solvers assessed the likelihood of success of contemplated approaches.

In Stage 2 of the model, we propose that people monitor progress using a criterion derived from the problem's current state. If the current operator meets the criterion, then it will be applied. In the case of the standard nine-dot problem this would virtually ensure failure to solve, because the current operator (move selection under a maximization heuristic) meets the criterion up to a point at which it becomes impossible to solve the problem. In other problem variants, a failure to meet the criterion creates the necessary conditions under which the problem space can be changed prior to a point at which the problem becomes insoluble. Criterion failure allows a search for a new or modified operator.

The second issue raised by Kaplan and Simon (1990) concerns the recognition or formulation of better moves. We follow Ohlsson (1990) in proposing constraint relaxation as a fundamental mechanism for changing the problem space in order to generate novel moves. That is, once the conditions for seeking alternative problem spaces are triggered we assume that the participant searches and selects, from the array of constraints, those that may be relaxed. In the case of the nine-dot problem the essential constraint that must be relaxed is the one that defines lines in terms of adjacent dots. If this constraint has a very low recognition threshold compared to others, then it may never be spontaneously relaxed without hints or other experimenter-supplied cues. The empirical evidence presented below suggests that this may be the case. As previously indicated, one of the mechanisms which we propose may trigger the search for a new or modified operator is the anticipation of failure signaled when the current operator fails to meet the required criterion of progress. For example, if the best and only move available under the model's constraints was one that cancelled zero dots (e.g., when passing through previously cancelled dots only) then this might well cause a participant to reassess the situation if there were many dots remaining and few moves. If, on the other hand, there were few dots remaining and many moves, a null-valued move may not appear so punishing. This suggests that the value of a move may be assessed not simply by the number of dots it cancels, but by the number of dots cancelled relative to the number of dots and moves left. Assumption 5 describes a criterion based on these considerations. Assumptions 6 and 7 then allow the use of non-dot points as one type of emergent move.

Assumption 5. The criterion used to judge progress is the ratio of the number of remaining dots to the number of lines available to cancel them.

Assumption 6. An unmarked location adjacent to a dot may be considered as having a dot if the result is a move that has a higher value than any other currently available. This unmarked point is considered to have the same properties as a previously canceled dot.

Corollary 6a. An unmarked location can never be used to start or end a move (since from Assumption 6 such extensions can never result in a higher value than the basic move itself).

Corollary 6b. Moves involving line extensions to an unmarked inflection point require a lookahead of more than one.

Assumption 7. When assumption 6 is invoked to continue a straight line across a gap then it is considered to result in an increased value for that line, and therefore implies no lookahead beyond the current line. In other words, an in-line extension across an unmarked point can be made with a lookahead of only one: without this assumption, the line would remain at the dot immediately before the gap.

To see how the model operates with these additional assumptions consider the twelve-dot problem (12A) shown in Figure 4. The problem can be solved as before, but the solution requires extending lines to a non-dot point. The problem is also shown in numerical notation in the left lower panel of Figure 4. The bracketed number (1) indicates the “missing” point.

Since solving the problem involves extending a line to the unmarked point (1) it follows from Corollary 6b that the problem cannot be solved with a lookahead of one line. With a lookahead of two lines, the problem is solvable. Under the previous rules, the best two-line first moves have a value of six. However, extending the line 13-5 to (1) and inflecting to the point 10 or 11 results in moves worth 7. Both of these moves are on the solution path. Two other moves of the same value are also available, 11-(1)-13 and 10-(1)-13, both wrong. The probability of a correct first move with two lookahead is therefore 0.5.

The impulse to seek emergent moves provided in assumption 6 might be considered to be either on-off or variable. In the former case, failure of the optimal move within the rules to meet the criterion would signal a search for alternatives. In the latter, the “strength” of the impulse to look for emergent moves might vary directly with the degree of discrepancy between the criterion and the current best move. Pursuing this second course, the impulse strength, m, might be considered proportional to $(c-b)/c$, where c is the value of the criterion and b the value of the best available move. If m is defined to have a minimum value of zero, then it can take on values from 0 to 1. However, while the likelihood of finding a correct emergent move may be proportional to the impulse to seek one, it will not necessarily equal that impulse, but may be influenced by other factors. To allow for this a parameter value, k, can be added, to give the equation $m=k(c - b)/c$. The value of the parameter may vary depending on factors such as instructions, cognitive load, time pressure, etc. Its importance is in accounting for effects of motivational and situational differences across participant groups.

This motivational component could then be incorporated into the calculations for problems involving non-dot points. Above, the likelihood of choosing a move involving a non-dot point was determined simply by the availability of such moves. The proposal is that this likelihood should now be multiplied by the impulse value, m. The probability of solving should then be equal to the product.

To illustrate, consider once again, problem 12A, again using a lookahead of two lines. The problem consists of 12 dots to be canceled in two moves, therefore the criterion for the first move is 6. The best moves available within the model's rules are also 6. Therefore, the motivation to seek emergent moves at this point is zero, and the operator will select one of the four moves with a value of 6. These are the moves 2-10-11, 4-11-10, 11-10-2 or 10-11-4. The first two are on a correct path while the last two are incorrect. The probability of selecting a correct move is therefore .5. Following the fate of the first correct move, 2-10-11, the best available second move has a value of 3 (the move 11-4-5), while the criterion is 6. The motivation to seek an alternative is therefore equal to k(6-3)/6, or .5k. This allows for the possibility that the solver will entertain extending a line to the non-dot location at (1), which makes available a correct move, 11-(1)-13, and no incorrect moves. The same is the case for the first move 4-11-10. Consequently, the conditional probability of finding this move given a correct first move will be 1.00x.5k, and the probability of a correct solution using a lookahead value of two will be proportional to .5x1x.5k.

The experiments

In the remainder of the article we report five experiments designed to test predictions from the model and to eliminate alternative hypotheses. The first experiment employed four conditions, one that used the standard nine-dot problem, and three that used different variants of the nine-dot, one each of eleven, twelve and thirteen dots. All four problems could be correctly solved by drawing exactly the same four lines. However, none of the three variants' solutions required extending lines either beyond implied figural boundaries or to non-dot points. The experiment tested a prediction derived from the model of a linear trend in performance across the nine, eleven, twelve and thirteen dot variants, reflecting the increasing ease with which early moves that lie on a correct solution path are found using a maximization heuristic. The results of the experiment supported the proposed model, but could also be partially explained because of differences in the cueing of starting points: Experiment 2 attempted to control for such possible effects. The experiment used the thirteen dot problem from the first experiment and two new variants, also with thirteen dots. All three problems had solutions that contained the same solution lines as the standard nine-dot problem, but varied in the salience of starting points. The results again supported the present model. The third experiment used three new twelve-dot problems created by omitting one dot each from three problems taken from the previous experiments. In each case the correct solution required extending lines to non-dot points. The experiment tested a prediction that making line extensions to non-dot points is a function of the impulse to seek alternative operators which arises from the early violation of a progress-monitoring criterion. The results supported the second stage of the model. Experiments 4 and 5 tested the second stage of the model on the nine dot problem itself, using the partial solution method of Weisberg and Alba (1981a) in which participants were given one of the alternative first lines of a solution as a hint. These experiments confirmed the model's prediction that, because a first line remaining within the implied figural boundary generates an immediate criterion failure, it will facilitate performance more than a first line which violates the implied figural boundary but that does not generate a criterion failure.

Experiment 1

Experiment 1 used four conditions: the nine-dot problem; the eleven-dot variant used by Burnham & Davis (1969); and a twelve- and a thirteen-dot variant created by adding extra dots on the diagonal of the eleven-dot version (see Figure 2). The experiment was conducted for two main reasons. First, it provided an empirical assessment of how much difficulty is contributed by having to extend lines and turn on non-dot points. This allowed an investigation of whether removing these sources of difficulty largely removes the problem difficulty. Second, it provided a test of Stage 1 of the present model.

The model’s predictions are derived from the kind of analysis shown above for the eleven-dot problem. Details of this analysis for the nine-, eleven-, twelve- and thirteen-dot versions are provided in the Appendix. A summary of the predictions is given in Table 1. For the nine-dot problem, Stage 1 of the model predicts the probability of success to be zero at lookahead values of one, two and three, and consequently few, if any, participants are expected to solve the problem. The three extra dot conditions are predicted to be substantially easier than this, with all or almost all participants expected to solve within ten attempts, requiring on average a mean of only a few attempts to succeed. Nevertheless, the model predicts differences between the three extra dot conditions, depending on lookahead values. Generally, the more dots added to the standard problem, the easier it should be.

Method

Participants. One hundred and thirty-four university undergraduates volunteered to participate. Of these, 22 were not naïve to the nine-dot problem and their responses were not included in subsequent analyses. Of the remaining 112, 27 attempted the nine-dot problem, 30 the eleven-dot, 25 the twelve-dot, and 30 the thirteen-dot. Responses were collected without identifiers; hence gender ratio and mean age of the participant group are not known.

Materials. The four different problems (nine-dot, eleven-dot, twelve-dot and thirteen-dot) are represented in Figure 2. Problems were laser printed in black on white paper. Dots were filled, 4.5mm diameter circles, arranged on a regular grid with 29.5mm between the centers of adjacent dots in both horizontal and vertical directions.

Each participant received a booklet of problems. On each of the first ten pages, one of the four problems was repeated in the same orientation. A cover page contained printed instructions, and four additional pages at the end of the booklet contained three repeated nine-dot problems followed by questions about whether the participant had seen the nine-dot problem before.

Procedure. The experiment was conducted during a lecture class. Participating students were randomly assigned to one of the four different problem conditions and given an appropriate booklet. Participants were first asked to read the cover sheet of the problem booklet, where the following standard instructions were given: “Your task is to connect the dots by drawing four connected straight lines, without either lifting the pencil from the page or retracing a line. For a solution to be correct the following must be satisfied: (1) every dot must be connected by a line; (2) you must use no more than four lines; (3) the lines must be straight; (4) once you have started drawing the first line you should not have to lift your pen from the page in order to complete the four lines. Participants indicated with a cross the starting point of solution attempts, to remain silent during the experiment, and to avoid collaborating with neighbors or looking at other people's solution attempts. Participants were allowed one minute for each attempt, and were requested to begin each attempt only when told. At the end of the experiment, the aims and purposes of the research were explained.

Results

Table 2 shows the total number of naïve participants in each of the four conditions, the percentage solving on the first trial, the percentage solving in 10 trials, and the average number of trials to solve (for those who solved within 10 trials). Since sample sizes were large and there were approximately equal numbers of participants in each cell, results for all three dependent variables were analyzed using analysis of variance. All statistical tests reported below used an alpha level of .05.

Overall performance

After one trial the percentage of solvers was 0%, 50%, 60% and 73% for the nine-, eleven-, twelve- and thirteen-dot conditions, respectively. Overall, there was a significant difference among conditions, $F(3,108)=15.78$, $Mse=.18$, $p < .001$. A trend analysis revealed a significant linear trend, $F(1, 108)=45.40$, $p < .001$, indicating that the proportion of participants solving on the first trial increased linearly with increasing number of dots. There was no significant departure from linearity. Overall, 61% of participants in the extra- dots conditions solved in their first attempt.

At the end of 10 trials the percentage of solvers was 0%, 93%, 80%, and 90% for the nine-, eleven-, twelve- and thirteen-dot conditions, respectively. There was a significant difference among conditions, $F(3,108)=68.12$, $Mse=.08$, $p < .001$. Planned comparisons revealed that all three extra dots conditions were significantly different from the nine-dot condition (all p's < .001), but not from each other. Overall, 88% of participants in these three conditions solved within 10 trials.

Of those who solved, the mean number of attempts to solution were 1.86, 1.38 and 1.22, for the eleven-, twelve- and thirteen-dot problems, respectively. There was a significant difference among the means, $F(2,73)=3.58$, $Mse=.81$, $p < .05$. Planned comparisons indicated that only the difference between the eleven-dot and thirteen-dot problems was significant ($p < .05$). Overall, the mean number of trials to solution for those who solved was 1.50.

Consistency with model

Solution attempts were classified as to whether or not they were consistent with the model. Any attempt or partial attempt that could be generated by the model's rules (operators at any of the four levels of lookahead) was classified as consistent, whereas those that could not be generated by the model were classed as inconsistent. Ambiguous or unclear attempts were left unclassified. Of the total of 112 first attempts, 107 were unambiguously classifiable, and of these, 85% were entirely consistent with the model and 90% followed the model's maximization principle. Of the 16 responses not consistent with the model, 11 used a shorter line than the maximum available, one connected non-adjacent points, and four required pencil-lifting. The proportion of inconsistent responses did not differ significantly by condition.

Performance on the nine-dot problem

The pattern of solution attempts across repeated trials were examined in detail for the standard nine-dot condition . The 27 participants were given up to 10 successive attempts to solve, and provided an average of 3.4 attempts. Overall, there were 88 classifiable responses. Of these, 60 (68%) were consistent with the model. Of the remaining 28, 23 involved either non-maximal lines or adjacent point violations or both, and four involved line extensions to non-dot points beyond the boundary of the dots. There were two cases of pencil lifting. Three (12%) of the 26 participants providing classifiable responses extended lines to non-dot points, indicating that this operator may emerge spontaneously in the case of the standard nine-dot problem. The evidence also indicated that participants were more likely to follow the model's rules on earlier than on later attempts. On the first attempt, 96% of responses were consistent with the model, on the second, 71%, while on subsequent attempts the proportion dropped to 53%. This suggests that repeated failures may lead to a search for other heuristics. However, the pattern of results did not suggest that there was any permanent abandoning of the primary heuristic. Of those making more than one attempt, a majority (76%) were still trying model solutions on their last attempt. In addition, of the 15 participants who made more than one attempt and who violated the model in at least one of those attempts, 7 of them (47%) made subsequent attempts that fully adhered to the model. Of the three participants who extended lines to non-dot points all three returned to making “model” moves in subsequent attempts.

Starting points

Participants were asked to mark the location where they began each solution attempt. Table 3 shows the percentage of participants starting from each location, the percentage solving from that location (in brackets), and the number reporting a starting point (the results are for the first attempt, since the majority of participants solved in one attempt). It can be seen that all three of the extra dots conditions were associated with a preference for starting from the bottom right hand location. Of the 93 participants who reported a starting location, 68% in the thirteen- dot condition started here, 55% in the twelve- dot and 58% in the eleven dot. In contrast, no one in the 9 dot condition reported this as a starting point. The overall difference was significant, χ^2 (3, N=93) = 25.24, $p < .001$. All three of the extra dots conditions differed significantly from the 9 dot in this starting preference, but none from each other. A second pattern suggested in Table 3 is the association between starting from the bottom right location and successfully solving a problem. For the extra dots conditions, 93% of participants starting from here solved in their first attempt, compared with 32% who started from other locations, χ^2(1,N=72)=14.01, $p < .001$. Of course, it is impossible to solve a problem from most of the other locations. However, the difference in success rates continued to hold when the bottom right location was compared with the top left, another point from which it is possible to successfully solve. For the latter location, the percentage of solutions was 56%, significantly lower than for those starting from the bottom right (Fisher's Exact test, N=60).

Discussion

Performance on the standard nine-dot task was at floor, as found in previous research. On the other hand, significant facilitation was found with the eleven-, twelve- and thirteen-dot variants. Importantly, on the first trial, three quarters of participants solved the thirteen-dot problem and half the eleven-dot problem, with performance on the twelve-dot problem lying intermediate between them. While adding the additional dots substantially ameliorated the difficulty of the nine-dot problem by eliminating the requirements either to draw lines outside of the figure's boundary or to start or finish a line on a non-dot point, it is clear that additional sources of difficulty must remain. This residual variation in difficulty across the problem variants can be explained by the present model.

The model was used to generate predictions concerning the probability of success on the first attempt, the probability of succeeding in 10 attempts and the mean number of attempts to succeed. There are several limitations in evaluating the model's predictions in a completely precise way. One is that the model's predictions vary with the hypothesized "lookahead" and we have no information concerning the lookahead capacities or strategies of the participants. A second is that we have no means of ascertaining whether participants tested possible solutions mentally before starting to sketch solutions, and consequently equating each page of the problem booklet with one solution attempt may err consistently in underestimating the number of attempts actually required to solve. For these reasons, the experiment may provide a more accurate test of the ordinal predictions of the model than of the exact values.

In terms of probability of success on the first attempt, the model, along with all other theoretical accounts, predicts that all three extra dot conditions will be superior to the nine-dot problem at all four of the possible lookahead types. This prediction was supported. The model also predicted that the twelve- and thirteen- dot problems would be easier than the eleven-dot for all but the maximum lookahead level, while the thirteen- dot would be easier than the twelve-dot for the first two lookahead capacities. Predictions concerning differences between the extra dot conditions are unique to the model, and are consistent with the findings of a significant linear trend across conditions and a significant superiority of the thirteen-dot condition over the eleven-dot.

As mentioned above, the fact that we do not know what mix of lookahead operators was present in the sample of participants makes it problematic to evaluate the more precise predictions of the model. For instance, for the eleven-dot problem, the predicted percentage that will solve on the first attempt ranges from 17%, if everyone uses a lookahead of 1, to 100%, if everyone has a lookahead of 4. Nevertheless, the observed percentage of correct solutions following the "first" attempt, of 50%, 60% and 73% for the eleven-, twelve- and thirteen-dot problems were in the ranges predicted by the model, of 17%-100%, 25%-100% and 50%-100%, for the three problems respectively. (See the rows labeled "probability of success" in Table 1 for the model's predictions.)

In terms of the number of participants successfully solving in ten attempts, the model's predictions were only partially supported. The model predicted success rates ranging between 84% and 100% for the 11-dot, between 94% and 100% for the 12-dot, and 100% for the 13-dot. The obtained values were 90%, 80% and 93%, respectively. While the results for the 11 and 13-dot conditions fell within the predicted ranges, those for the 12-dot did not. No significant differences were found between the three extra dot conditions..

With respect to the mean number of attempts required to solve, the model predicts that the thirteen- and twelve-dot versions should be simpler than the eleven-dot for all lookahead capacities except the maximum, while the thirteen- dot will be easier than the twelve for the first two lookahead levels. The finding that significantly fewer mean trials were required to solve the thirteen- dot than the eleven is consistent with these predictions. The observed values of 1.8, 1.4 and 1.2 for the eleven-, twelve- and thirteen-dot conditions- again fell within the ranges predicted by the model, of 6 to 1, 4 to 1 and 2 to 1 for the three problems respectively. However, in each case the observed values fell towards the low end of the predicted ranges, suggesting that participants were solving faster than predicted by the model. There are a number of possible factors that may have contributed to this. First, the observed mean was calculated using only the data from participants who solved within ten attempts, whereas the model's predictions are based on the sum of a series that extends beyond ten attempts, to infinity. Second, participants may have made more attempts than were recorded, if they mentally considered moves before sketching solutions. This would tend to lower the apparent number of attempts to succeed. Third, the model's simplifying assumptions that optimal moves are selected randomly with no preferences, and that moves are sampled with replacement across subsequent attempts (i.e. no learning occurs) may be too simple. A model that assumed that incorrect moves were sampled without replacement would predict lower mean trials to success.

In terms of the model's more detailed predictions, an analysis of the first solution attempts across the four conditions of Experiment 1 indicated substantial support. A large majority of solution attempts (85%) were consistent with the rules of the model, and even more (90%) were consistent with the model's main assumption, that moves are made that connect the maximum number of dots. A detailed analysis of solution attempts across successive trials of the standard nine-dot condition provided similar support. Of the 88 classifiable responses, 60 (68%) were consistent with the model. Of the remaining 28, 23 involved either non-maximal lines or adjacent point violations or both, and four involved line extensions to non-dot points beyond the boundary of the dots. There were two cases of pencil lifting. Three participants (12%) extended lines to non-dot points, indicating that this operator may emerge spontaneously even in the case of the standard nine-dot problem. Participants were more likely to follow the model's rules on earlier than on later attempts with 92% of first attempts consistent with the model, dropping to 53% for the third and subsequent attempts. This supports the suggestion that repeated failures may lead to a search for other heuristics. However, the pattern of results did not suggest that there was any sudden or dramatic abandonment of the primary heuristic. Rather, the evidence suggests that people experimented with minor modifications to the model, then returned to it, then tried modifications again. Of those making more than one attempt, a majority (76%) were still trying model solutions on their last attempt. Of the three participants who extended lines to non-dot points -- the "breakthrough" required to solve the problem --none solved it, and all three returned to making conventional moves in subsequent attempts. This does not suggest any sudden radical departure from one mode to another, but rather a series of faltering attempts to break free of a compelling heuristic. The model and evidence are also consistent with Weisberg's (1995) view that the nine-dot problem may be tackled through trial and error. However, the results do not support his suggestion that the problem may be *solved* in this way, in that no correct solutions appeared. According to the model, a solution to the nine-dot problem can be found only by applying a lookahead capacity of four lines. In other words, the solver would require enough capacity to be able to "see" the construction of the entire solution. This suggestion is consistent with the phenomenological accounts of insight that emphasize the emergence into consciousness of a more-or-less complete solution.

Experiment 2

In spite of the general support for the model's predictions, another possible explanation for the results of Experiment 1 exists: Adding extra dots to create the eleven-, twelve- and thirteen-dot conditions may have cued participants to start from the bottom right location. Since this is one of only two correct starting points for these problems, this alone might have led to the observed facilitation over the nine-dot problem. Furthermore, if starting point selection was constrained by the addition of extra dots to the diagonal tail of the twelve- and thirteen-dot problems, then this could explain the pattern of differences between the extra dot conditions. The model assumes equal preferences for moves that meet the model's rules. This assumption is not supported by the pattern of results for the nine-dot condition shown in Table 3, which shows that there are strong pre-existing starting preferences. Seventy-five percent of first attempts to solve the nine-dot problem started from either the top left or the bottom left. Pre-existing starting-tendencies may consequently have influenced choices in the other conditions also.

Because of this possibility, a second experiment was conducted where the prominence of starting-points was manipulated. Three problems were used (see Figure 3): Problem 13, the thirteen-dot problem of Experiment 1; Problem 13A, created by adding two dots at the top of the eleven-dot problem; and Problem 13B, created by adding one dot at the top of the twelve-dot problem. The problems were designed to manipulate prominence of starting-point in two somewhat different ways. Problems 13 and 13A are alike in that each contains an "outlying" dot two units beyond the main body of the figure. If this is the factor that signals a prominent starting point, then Problems 13 and 13A should be simpler than 13B, which has no comparable outlying dot. Alternatively, a salient starting-point may be one that lies at the end of the longest line of dots. If this is the case, then 13A should have the most prominent starting-point, lying at the head of a column of length six. This would make Problem 13A simpler than either 13 or 13B, which should be equally difficult, having maximum line lengths of five dots. The model, on the other hand, predicts a third pattern of outcomes -- no difference between 13A and 13B, with 13 being simpler than both (at 1 and 3 lookahead values). The model's predicted probability of solving on the first attempt for lookahead values of one, two, three and four respectively are: for Problem 13, .5, .5, 5 and 1: for Problems 13A and 13B, .25, .5, .33 and 1. The underlying analysis is given in the Appendix.

Method

Participants. Ninety-seven university students volunteered to participate. Responses from 20 who reported seeing the nine-dot problem previously were not analyzed further. Of the 77 remaining, 25 attempted Problem 13, 27 Problem 13A, and 25 Problem 13B. As in Experiment 1, the gender ratio and mean age of the participant group are not known.

Materials. The three problems are shown in Figure 3. Problems were printed using the same dot and grid parameters as for Experiment 1, except that the horizontal and vertical distances between dot centers were 27.5mm. Problem booklets comprised three pages: the first containing instructions, the second containing one of the problems (with an instruction to indicate the starting point of the solution attempt), and the third containing the nine-dot problem together with questions on whether this problem was familiar to the participant.

Procedure. The experiment was conducted as part of a lecture class on problem-solving. Participating students were randomly assigned to one of three problem conditions and given an appropriate booklet. Instructions to participants were similar to those given in Experiment 1. Two minutes were allowed for each of the two problems.

Results

There were 27 naïve participants in condition 13B and 25 in each of the other two conditions. The percentages of participants solving in one trial were 88%, 55.6%, and 56%

for conditions 13, 13A, and 13B, respectively. There was a significant difference among conditions, $\underline{F}(2,74)=4.20$. $\underline{MSe}=$ $\underline{p}=.019$. Planned contrasts showed that 13 was significantly easier than both 13A, (p=.008) and 13B ($\underline{p}=.012$), while 13A and 13B did not differ.

Of 77 responses, 72 were classifiable as consistent or inconsistent with the model. Of these, 54 (75%) were consistent. Of the 18 inconsistent responses six contravened both the maximization and adjacency rules, six the maximization rule, four the adjacency rule, while the remaining two required pencil-lifting. The proportion of inconsistent responses did not vary significantly by condition.

Table 4 shows the percentage of participants reporting various starting points. The percentage solving from each location is shown in brackets. As in the first experiment, the majority of the 23 participants in condition 13 who reported a starting point started from the bottom right location (74%). In condition 13A, 74% of the 23 reporting started from the top left location. Similarly, for condition 13B, 75% of the 20 reporting started from the top left. For these conditions, the salience of the preferred starting-point was virtually identical.

Discussion

The results of Experiment 2 were consistent with the predictions of the model, in that performance on Problem 13 was significantly better than on both of the other problems, neither of which differed from each other. Starting-preferences could not explain the differences between conditions. In fact, the percentage of participants selecting the most popular starting-point was essentially the same across the three conditions.

While Experiments 1 and 2 support the present model, none of the problems they used required turning on non-dot points or extending lines beyond the boundaries of the given shape. Rare as moves of this kind are in naïve attempts to solve the nine-dot problem, they do occur. Stage 2 of the model covers moves of this kind, and generates predictions that are tested in Experiments 3, 4 and 5. Before describing Experiment 3, we return to the concept of lookahead. In particular, we explore an approach to estimating the mix of lookahead values, using the data from Experiments 1 and 2. This is feasible since performance in these experiments is essentially based on Stage 1 of the model, and therefore data are not complicated by a requirement to find an alternative operator. As such, they allow a relatively pure estimate of lookahead, which can be used to derive more detailed predictions from Stage 2 of the model, tested in subsequent experiments.

Estimating Lookahead

The results of Experiments 1 and 2 were used to calculate the relative proportion of different lookahead types that provided the best fit between model and data. This was done by finding the mix of lookahead values that maximized the coefficient of determination between observed and predicted percentage of participants solving on the first attempt. All possible combinations of lookahead percentages from 0 to 100 were used for each of the four lookahead types, with the constraint that the four had to sum to 100%. To make the calculation feasible only integer values were employed. In addition, the value of the parameter k was varied from 0 to 1 in increments of 0.01.

With no additional constraints the best fit was found when the mix was 60%, 0%, 16% and 24% for lookahead types one through four, respectively (r^2=.97). This was obtained at a parameter value of k=0.96. Since this was very close to unity, and since setting k=1 made virtually no difference to the goodness of fit, with an r^2 value equal to the maximum to the fifth decimal place, a value of k=1 was adopted. Doing so had no effect on the best mix of lookahead values, which remained at 60%, 0%, 16% and 24% for lookahead types one through four. However, although technically providing the best fit, there are a number of reasons to treat this estimate with caution. Importantly, it was only one of 32000 quite varied solutions that provided an r^2 value greater than .96 (with k=1). Clearly these are all well within sampling error of each other. Another concern was that it estimated the proportion of two lookahead users to be zero, which seems unlikely to be the case. Further, it seems unreasonable that the distribution of types should be U-shaped going from one to four. Rather, it seems more likely that the distribution would be unimodal. Finally, since four lookahead would require a processing capacity well beyond the normal upper limit, it seems extremely unlikely that one quarter of the population would operate at this level. Consider that at two, three and four lookahead each successive move in the 9-dot problem is essentially a binary choice (since at the end of each move there are two maximum moves available for the next). Operating at four lookahead would therefore require considering sixteen alternatives, well beyond the normal capacity of 7 plus or minus 2. This suggests that the percentage operating at four lookahead would be extremely low, perhaps close to zero. This is consistent with the fact that the nine-dot problem is very rarely solved, even though at four-lookahead, the solution is within reach, according to our model.

In order to reduce this array of 32000 solutions and respond to some of the other concerns we repeated the exercise with two additional constraints --that the distribution should be unimodal and that the percentage of four lookahead types be zero. These constraints severely limited the set of solutions with r^2 values as high as .96 to just 12. For all 12 of these solutions the mix of lookahead types from one to three was roughly equal. The estimates ranged from 30% to 36% for one lookahead, from 30% to 33% for two lookahead and from 32% to 40% for 3 lookahead. The best fit was provided at 32%,32% and 36% for one through three lookahead, respectively (r^2=.96, $p<.01$). At first sight, a rectangular distribution across one, two and three lookahead may seem surprising; however, if all three of these lookahead levels are within the limits of normal processing capacity, any variation may have more to do with motivation or level of effort than with capacity, in which case, a flat distribution may be no more unlikely than any other specific distribution.

Clearly, these estimates should be used with caution, and there may be some concern that constraining the proportion of four lookahead operators to 0 may be too extreme. We comment further on the general issue of estimating lookahead in the General Discussion However, in each of the experiments, the predicted outcomes at four lookahead are identical across experimental conditions, meaning that four lookahead essentially adds a constant to the predictions, and eliminating it is not likely to affect the comparison between predicted and obtained results.. In subsequent experiments the estimated proportions of 32%, 32%, 36%and 0% for lookahead types one through four were used to derive model predictions.

<u>Experiment 3</u>

In the third experiment the 13 and 13A problems of Experiment 2 were modified to create problem displays in which solutions required non-dot turning points (see Figure 4).

The summary predictions of the model are as follows, for lookahead values of one, two, three and four, respectively: for Problem 12A, 0, .25k, .056k and .167k (where k is the parameter value); for Problems 12B and 12C, 0, .125k, 0 and .083k. The underlying analysis is given in the Appendix. Using the parameter estimates derived above, the following more specific predictions result. At the end of one trial, the percentages solving should be 10.0%, 4.0% and 4.0% for Problems 12A through 12C, respectively. The corresponding percentages at the end of four trials will be 29.3%, 13.2% and 13.2%. The analysis indicates that the percentage solving should be greater for Problem 12A than for 12B and 12C, which should not differ from each other. On the other hand, if turning on non-dot points were the sole source of difficulty in problems of this kind, these three variants should be equally difficult.

Method

Participants. Two hundred and twenty university and college students volunteered to participate. Twenty-seven were not naïve to the nine-dot problem, and their responses were not considered further. Of the 193 remaining, 63 attempted Problem 12A, and 65 each attempted Problems 12B and 12C. The gender ratio and mean age of the participant group are not known, as in Experiments 1 and 2.

Materials. Problems were printed using the dot and grid parameters of Experiment 1, except that horizontal and vertical distances between dot centers were 25.5mm. Each booklet contained seven pages: the first giving instructions, followed by one of the twelve-dot problems repeated on each of four pages, followed by two pages giving the nine-dot problem and questions about its familiarity.

Procedure. The experiment was conducted in three separate lecture and seminar classes. Participating students were randomly assigned to problem condition and issued with an appropriate booklet. Instructions were similar to those given in the previous experiments. One minute was allowed for each attempt at a problem. At the end of the experiment, participants were verbally informed about the purpose of the research.

Results

There were 63 participants in condition 12A and 65 each in conditions 12B and 12C. The numbers (percentages) solving on the first trial were 7 (11.1%), 4 (6.2%) and 4 (6.2%) for the three conditions respectively. The corresponding predicted values were 6.3 (10.0%), 2.6 (4.0%) and 2.6 (4.0%). The numbers (percentages) solving within four trials were 17 (27.0%), 9 (13.8%) and 7 (10.8%), and the corresponding predicted values, 18.5 (29.3%), 8.6 (13.2%) and 8.6 (13.2%). There were no significant differences in performance at the end of one trial, F(2,190), = .72, MSe=.07, p=.49. This is in line with the model predictions, where the expected results using these sample sizes give F=.87, MSe=.06, p =.42. There was a significant difference after four trials, F(2,190)=3.40, Mse=.14, p < .035. Planned comparisons indicated that condition 12A was significantly superior to the other two conditions (p < .05), which did not differ from each other. Again, the result was consistent with the predicted outcome, which yielded F=3.08, MSe=.15, p=.048.

The goodness-of-fit of predicted to obtained proportions of success was tested using the Kolmogorov-Smirnov procedure. Predicted and obtained frequencies did not differ significantly after one trial (DMAX=.08) or after four (DMAX=.02). Of those who solved, the mean number of trials to success (and standard deviations) were 2.12 (1.17), 1.89 (1.05) and 2.00 (1.41) for conditions 12A through 12C, respectively. There were no significant differences among conditions, F(2,30)=.11. This absence of difference is consistent with the model, with corresponding predicted means (standard deviations) of 2.22 (1.13), 2.33 (1.18) and 2.33 (1.18), giving F=.04 for these sample sizes. (Note that the predictions for mean trials were calculated from the predicted percentages solving on Trials 1, 2, 3 and 4, not from the general formula for Bernoulli trials, since the latter is based on an infinite series of trials.)

Attempted solutions (first attempts) were examined for inconsistencies with the model (including turning on non-dot points). Of the 193 first attempts, 144 (75%) were unambiguously classifiable. Of these, 61 (42%) were consistent with the Stage 1 model. The percentage of consistent responses was 33%, 57% and 36% for the 12A, 12B and 12C conditions, respectively. There was a significant difference among conditions, F(2,141)=3.63, Mse=.24, p=.03, with the 12A and 12C conditions showing significantly less consistency than the 12B. The 83 first attempts that were inconsistent with the Stage 1 model exhibited the following departures from 'model' behavior: 52 had instances of failure to use the longest available line (40 "pure" cases and 12 in combination with other deviations), 5 connected non-adjacent dots, 2 required pencil- lifting, while 24 involved line extensions to non-dot points. The frequency of starting points selected by participants on their first solution attempt is shown in Table 5.

Discussion

As would be expected by all of the theories, performance with the twelve-dot variants of Experiment 3 was reduced relative to the thirteen-dot problems of Experiment 2. One might, however, question how participants were able to solve the problems at all if effects of specific experience or constraint imposition were the sole factors determining performance with these problems. Neither Weisberg & Alba (1981a) nor Lung & Dominowski (1985) report trial by trial data for their experiments in which nine-dot performance was facilitated, so it is not easy to compare the relative levels of performance across experiments. Nonetheless, the solution rate observed here with Problem 12A, of 27% after just four one-minute trials does seem to demonstrate some degree of competence with the problem that would not be predicted by the specific experience or constraint imposition accounts. By way of comparison, Lung and Dominowski (1985) found that 22% of participants given practice problems were able to solve the nine-dot problem after 20 attempts.

More compelling evidence from Experiment 3 concerning the limited roles of specific experience and constraint imposition comes from the observed facilitation with problem 12A relative to the other twelve-dot problems. Solution rates for this condition were roughly twice those of the others, both after one attempt and at the end of four. This pattern of findings is predicted by the present model and is incompatible with previous accounts.

Although the results of Experiments 1, 2 and 3 are consistent with the model, and generally run counter to the predictions of explanations based upon hypothesis-testing (Weisberg and Alba, 1981a) and constraint-imposition (Lung and Dominowski, 1985), there remains the possibility that the effects reported so far are a result of figural integrity. For example, the increasing facilitation over 11-, 12- and 13-dot problem variants in Experiment 1 might result from the increasing goodness of figure, or perhaps more specifically the increasing recognisability of an 'arrow-like' figure in the problem array. The figural properties of the array might make available the arrow shape to serve as a model for a putative solution. Similarly, in Experiment 2, it might be that the 13-dot problem simply offers a more recognizable figure than the 13A and 13B problem variants. Finally, the greater facilitation found with problem 12A in Experiment 3 may simply result from the recognition of an arrow shape, albeit incomplete, in the problem array. Experiment 4 was conducted in order to provide a critical test between the Gestalt account of figural integrity and the current model, in which performance on the standard 9-dot problem was investigated using the partial solution method of Weisberg and Alba (1981a). The model generates predictions that are in precisely the opposite direction to those deriving from Gestalt, hypothesis-testing and constraint-imposition accounts.

Experiment 4

In their first experiment, Weisberg and Alba (1981a) used four conditions, where participants were initially given 10 attempts at the standard nine-dot problem. In the first condition (control) those who failed to solve by this point were then given an additional 10 attempts. In each of the other three conditions, participants who had failed after 10 trials were told that they had exhausted all solution possibilities within the figure and that they should move outside (the 'exhausted' hint), followed by 10 additional trials. The second condition used the standard nine-dot problem following the exhausted hint. The third condition used a version with a first correct solution line already drawn. This was a diagonal line that started from outside the array of dots and extended in a straight line through dot 9 to dot 1 in Figure 5. The fourth condition gave the problem with the first two lines correctly drawn. The percentages of participants solving the problem within the second set of 10 trials were 0%, 20%, 62% and 100%, for the four conditions, respectively. All three experimental conditions produced a significantly higher percentage of correct solutions than the control. In addition, the fourth condition was significantly better than all others.

Of particular interest here is the fact that giving a first correct line produced a significant and substantial facilitation in performance over the control condition. What is not entirely clear is why this occurred, in that at least three potential factors may have played a role. The first is the exhausted hint, itself. The second is the fact that the given line extended beyond the array of dots, providing a visual example of a line extension to a non-dot point. The third, arising from the model proposed here, is the fact that giving the first line means that the problem can be solved with a minimum lookahead of three, rather than of four. In addition, providing a first line changes the availability of model moves that meet the criterion, and therefore alters the impulse to seek emergent moves. According to Weisberg and Alba (1981a) the exhausted hint by itself was of little help, because participants possessed no task-specific knowledge that they could apply to the problem so that, when told to "go outside", they had no idea how to proceed. Experiment 4 was designed to clarify some of these issues and, more specifically, to test predictions from the present model in the context of the nine-dot problem itself.

The experiment used three conditions. All three followed the general format of the Weisberg and Alba experimental conditions described above, in that participants were given 10 trials on the standard form of the nine-dot problem, followed by the exhausted hint, followed by 10 trials under experimental conditions. All three experimental conditions presented the nine-dot problem with a first correct line provided. The three conditions are illustrated in Figure 5, as 9A, 9B and 9C. Two of the conditions (9A and 9B) provided a visual indication that lines can be extended to non-dot points beyond the implied boundary of the dot-array. The third (9C) did not. Theoretical explanations that emphasize the role of line extensions to non-dot points (Lung and Dominowski, 1985), or that emphasize the role of the figural boundary, (Sheerer, 1963) would presumably predict that the first two conditions should facilitate performance over that of the third condition, which has no line extension. In addition, it seems reasonable to expect that condition 9A, where the line extends to a non-dot point that is on a correct solution path, should provide a more useful visual hint than in 9B, where the line extends to an irrelevant point. Indeed, 9A may provide just the kind of task-specific knowledge that Weisberg and Alba (1981a) observed to be lacking with the exhausted hint alone. It appears, then, that previous explanations predict an ordered outcome, in terms of the figures' capacities to facilitate correct solutions, that may be summarized as 9A > 9B > 9C. The key prediction is that figures 9A and 9B should be superior to 9C.

In contrast, the model predicts 9B=9C>9A. The model's predictions arise in the following way. (Details are in the Appendix.) With a lookahead of one, the criterion is two dots, and the best move(s) have a value of two for all three figures. Consequently, there is no impulse to look for an emergent move, resulting in failure in all three conditions.

With a lookahead of two, differences emerge between conditions. The criterion for two lookahead is four dots. For condition 9A the best available model moves meet this criterion. One example is the move going from dot 1 through dot 7 to dot 9 (assuming that participants follow the instruction that they must start from one or other end of the first line). Thus, there is no reason to extend lines beyond dots in this case, resulting in failure. For conditions 9B and 9C, on the other hand, the best model moves have a value of three, resulting in a value of .25k for the impulse to seek moves beyond the model's heuristic. This makes it possible for a participant to seek and find a move that satisfies the criterion. One move that does so is shown in Figure 6, where the first line through points 1,4 and 7 is extended to a non-dot point, allowing the second line to pass through points 8 and finish at point 6, thereby meeting the criterion of canceling four dots. There is no corresponding need to extend a line to a non-dot point for the horizontal condition, leading to the prediction that emergent moves that extend lines will appear on the first test trial in the diagonal line conditions, but not in the horizontal.

At two lookahead, making the line extension shown in Figure 6 should not lead to success on the first trial, since there would be no reason to make the required second line extension (beyond point 6). Nevertheless, making the initial line extension move may form the basis for success in subsequent trials. We use the term "promising state" to refer to precisely this situation: where a novel move satisfies the local criterion operating under a relaxed constraint (specifically, model definition 2) at a particular lookahead value, but the problem has not yet been solved. It is "promising" because the move has value (in meeting the criterion) and may therefore be more likely to be remembered and repeated in later attempts. In fact, the two lines that comprise it may effectively form a chunk which, in subsequent attempts, consume only one unit of lookahead rather than two. This would permit solution of the problem with a lookahead of two, rather than of three. In summary, for the diagonal line condition with a two-lookahead operator, the conditions for insight are provided by the impulse to seek emergent moves of .25k, whereas a promising state consisting of the move 1 - (non dot point adjacent to 7) - 6 may represent the source of insight required to make a further line extension and thus solve the problem. We therefore expect no differences between conditions in the numbers of two-lookahead users solving on the first test trial, as sources of insight have not yet been recognized, but we anticipate that differences will occur in subsequent trials, in favor of the diagonal line conditions.

At three lookahead, the criterion is six dots, and no model moves meet this criterion for any of the three conditions. The best moves available for all the figures have a value of 5, resulting in an impulse *strength* of m=.17k. Consequently, the tendency to solve at three lookahead should be equal across conditions.

The model therefore predicts an ordering of 9B=9C >9A. Only a weak ordering among conditions is predicted, since it emerges solely because of the difference in impulse strength, m, at two lookahead. Nonetheless, the fact that it is opposite to the ordering derived from other theoretical approaches provides for a critical test.

Using the previous estimates for lookahead types, the following more specific predictions are generated. The percentage of participants solving on the first test trial will be given by the percentage of three lookahead types multiplied by the probability of their solving, or 6% for each condition. At the end of ten trials, the percentage solving in the horizontal condition will be 30% (limited to those operating at three lookahead). The proportion solving in the diagonal conditions will equal this plus those operating at two lookahead who can capitalize on making the emergent move shown in Figure 6. If everyone at two lookahead did so, this would result in an additional 30% solving. For those who solve, the mean (and standard deviation) of the number of trials to solve is predicted to be 4.05 (2.77) in the horizontal condition. For the diagonal conditions, the predicted value may be somewhat less than this, but no lower than 3.45 (2.57), which would occur if all two lookaheads solve. Thus, we do not anticipate any substantial differences between conditions in the number of trials required for those who solve.

The fact that the model predicts greater facilitation from a correct first line remaining within the figural boundary (9C) than from one that extends beyond the boundary (9A) is a strong and counter-intuitive prediction. However, there is an alternative source of such a prediction, namely that the line given in figure 9C offers two starting points that can lead to a correct solution, whereas figure 9A offers only one. If participants tackle the nine-dot problem purely through a process of trial and error, a possibility suggested by Weisberg (1995), then the likelihood of finding a correct solution by this method will be greater in figure 9C than 9A. In order to address this possibility, figure 9B was included in the experiment. Like figure 9A, it offers only a single starting point that can lead to a correct solution. If the number of available starting points is the crucial determinant of performance, then the predicted ordering is 9C>9B=9A.

Method

Participants. 134 members of the general public were tested during demonstrations conducted as part of University Open Days. As in previous experiments, identifiers were not collected and hence the age and gender distributions of participants are not known.

Materials. The three different test problems (9A, 9B and 9C) are represented in Figure 5. Problems were laser printed in black on white paper. Dots were filled, 3mm diameter circles, arranged on a regular grid with 13.5mm between the centers of adjacent dots in both horizontal and vertical directions. Dot size and spacing was reduced from previous experiments simply in order to reduce booklet page size, for reasons of economy. Each participant received a booklet of problems. A cover page contained printed instructions, then on each of the subsequent ten pages, the standard 9-dot problem was presented. The next page contained an instruction to listen to further instructions, and was followed by ten pages each containing the 9-dot problem with one line superimposed, as shown in Figure 5. A final page asked whether the participant had seen the nine-dot problem before.

Procedure. Participants were assigned to conditions by random distribution of booklets at the point of entry to the lecture theatre. Participants were first asked to read instructions, presented on an overhead projector, for the standard 9-dot problem. These instructions were exactly as for Experiments 1 - 3. Participants were then allowed 30s for each of ten attempts at the standard 9-dot problem, and were requested to begin each attempt only when told. A new set of instructions were then presented by overhead projector, reading as follows: "The task in the next pages is identical, except that one line has been given to help you. Draw three additional straight lines, starting from one or other end of the given line, to connect all 9 dots. Hints: 1. Your lines can go outside the square shape formed by the 9 dots. 2. Your previous attempts have probably exhausted all the possibilities that exist within the square shape". A reminder of the requirements for a correct solution was then given. The relevant instructions were visible throughout each stage of the experiment.

Results

Twenty-four of the 134 participants indicated that they had seen the nine-dot problem previously, and were eliminated from further analysis. In addition, six of the remainder solved the standard nine-dot problem within the initial ten control trials, and were also eliminated from the overall analysis. Of the remaining 104 there were 36 in condition 9A and 34 each in conditions 9B and 9C.

Performance on the nine-dot problem

The number (percentage) of participants solving the problem on the first trial of the second set of 10 attempts was 1 (3%), 3 (9%) and 2 (6%) for conditions 9A through 9C, respectively. The number (percentage) solving within the second set of 10 attempts was 10 (28%), 16 (47%) and 16 (47%) for conditions 9A through 9C, respectively. The means (standard deviations) of the number of trials to solve were 4.30 (2.87), 3.75 (2.86) and 3.88 (2.60), respectively.

As predicted by the model, there were no significant differences between the two diagonal line conditions on any of these measures. Consequently these conditions were combined for the next two analyses, which therefore compared the horizontal line given condition with the diagonal. The difference between the two conditions in the number of participants solving on the first test trial was not significant, (t (102)=.95). The number (percentage) solving within the ten test trials was 10 (28%) and 32 (47%) for the horizontal and diagonal conditions, respectively. The difference was significant in the predicted direction, $t(78)=1.98$, $p=.025$ (separate variances test, one-tailed). The means (standard deviations) of the number of trials to solve were 4.30 (2.87) and 3.81 (2.69) for the horizontal and diagonal conditions respectively. The difference was not significant, t (40)=0.49.

Mechanisms underlying performance differences

The performance differences predicted by the model were based on differences in predicted levels of motivation to seek emergent moves, including moves that extended lines to non-dot points, a requirement of solution. The model allows for such moves to appear in first attempts, since it hypothesizes that they arise in an anticipatory way, from seeking moves to meet a current criterion, and not simply in reaction to the experience of a failed attempt. The number (percentage) of first solution attempts that extended lines to non-dot points were 8 (22%), 14 (41%) and 13 (38%), for conditions 9A through 9C, respectively. While in the predicted direction, the overall difference was not significant. Since again the results for the two diagonal conditions were almost identical, their results were combined to compare a horizontal with a single diagonal condition. The resulting difference was significant in the predicted direction, t(81.7)=1.90, p=.03 (separate variances test, one-tailed).

The number (percentage) of participants exhibiting the move shown in Figure 6 or a similar move was 5 (14%), 6 (18%) and 16 (47%) in conditions 9A through 9C, respectively. The overall difference was significant, F(1,101)=6.48, MSe=.175, p=.002. Contrasts showed the short diagonal condition to be significantly higher than both other conditions, which did not differ significantly.

Making this move was associated with eventual success. Of the 98 participants who did not solve on the first attempt and who therefore had an opportunity to exhibit the move, 27 participants did so. Of these, 23 (85%) eventually solved the problem. Of the 71 who did not, only 13 (18%) solved. In addition, the move was usually quickly followed by a successful solution. Of the 23 solvers who made the move, 13 (57%) solved on the trial immediately following. Of the 10 who didn't solve immediately after first making the move, 8 repeated it at least once before solving, suggesting that it was recognized as valuable.

Discussion

The results of Experiment 4 supported the model predictions. As predicted, there were no differences between conditions on the first test trial, with success rates of 3%, 9% and 6% for conditions 9A through 9C, compared with a predicted rate of 6% for all conditions.

At the end of 10 trials the success rate for the horizontal condition was 28%, close to the predicted rate of 30%. We had no correspondingly precise prediction for the diagonal conditions. The success rate could have been as low as 30% if no two-lookahead users capitalized on the emergent move. It could have been as high as 60%, if all did so. The obtained values of 47% for both diagonal conditions fell between these limits. While the success rate was higher for the diagonal conditions than the horizontal, there was no significant difference between conditions in the number of trials to solution for those who solved. This was anticipated from the model.

There was some evidence to support the notion of "promising states" underlying the performance differences between conditions. The idea was that criterion failure at two-lookahead in the diagonal conditions, and the subsequent relaxation of the constraint (definition 2) that limits lines to adjacent points, would allow emergent moves such as that illustrated in Figure 6 that meet the criterion. Such moves, if found, might be memorable (for example, via a process of chunking into a single move) and repeated. The results showed that such moves were more likely in the diagonal than in the horizontal conditions, but significantly so only for the short diagonal condition. The results also showed that the move was likely to lead to success in the immediately following trial. If not, it was likely to be repeated. Making this move was strongly associated with eventual success.

While we cannot claim that the results provide unequivocal support for the model, they seem nevertheless to be important. The overall result, that after 10 trials significantly more solutions occurred in the diagonal than the horizontal condition, seems to challenge current views. Although the effect size appears relatively small (which is consistent with the model), the fact that all other accounts would seem to predict the opposite ordering of problem difficulty offers, in our view, considerable support for the current model. The results suggest that providing a first line that goes outside the implied boundary of the problem array is, in and of itself, not sufficient to lead to insight into the problem's solution. Similarly, the number of available starting points that allow correct solutions to be produced does not appear to be the key determinant of performance, a point demonstrated by the absence of a difference in performance between figures 9B and 9C. Instead, it is necessary that the given line generates an early impulse to seek alternative operators through the failure to meet the progress monitoring criterion. This happens at two and three lookahead in the case of problems 9B and 9C but only at three lookahead in the case of problem 9A.

Experiment 5

Because of the potential importance of these findings, Experiment 5 attempted a replication. The experiment focussed upon the critical test offered by the comparison between conditions 9A and 9C of the previous experiment.

Method

Participants. One hundred and three university undergraduates were paid £3 each to participate. As in previous experiments, identifiers were not collected and hence the age and gender distributions of participants are not known.

Materials and Procedure. The materials and procedure for this experiment were identical to those of Experiment 4, with the omission of problem 9B.

Results

Sixteen of the 103 participants indicated that they had seen the nine-dot problem previously, and were eliminated from further analysis. In addition, four of the remainder solved the standard nine-dot problem within the initial ten control trials, and were also eliminated from the overall analysis. Of the remaining 83 there were 40 in the horizontal condition and 43 in the diagonal.

Performance on the nine-dot problem

The number (percentage) of participants solving the problem on the first trial of the second set of 10 attempts was 3 (8%) and 9 (21%) for the horizontal and diagonal conditions, respectively. The difference was not significant, (t (72.56)=1.78, p=.08, separate variances test). The number (percentage) solving within the second set of 10 attempts was 17 (43%) and 28 (65%) for the two conditions, respectively. The difference was significant in the predicted direction, $t(81)=2.10$, $p=.02$ (one-tailed). For those who solved, the means (standard deviations) of the number of trials to solve were 4.35 (2.67) and 3.39 (2.30), respectively. The difference was not significant, $t(43)=1.28$, $p=.21$.

Mechanisms underlying performance differences

The numbers (percentages) of participants exhibiting a move similar to that shown in Figure 6 were 10 (25%) and 22 (51%) in horizontal and diagonal conditions, respectively. The difference was significant, $t(80.62)=2.52$, $p=.01$, separate variances test. Again, making this move was associated with eventual success. Of the 32 participants making this move, 20 (63%) solved the problem, whereas of the 39 who did not, 13 (33%) solved, χ^2 (1)= 6.01, $p=.01$. In addition, the move was usually quickly followed by a successful solution. Of the 20 solvers who made the move, 7 (35%) solved on the trial immediately following. On average, these participants solved in two attempts of first making the move.

Discussion

The results of Experiment 5 replicate those of Experiment 4, and provide further support for the model predictions. One difference with Experiment 4 concerns the absolute level of performance, with roughly 20% higher solution rates in Experiment 5 than in Experiment 4. The solution rates of Experiment 5 are somewhat closer to those found by Weisberg & Alba (1981a, Experiment 1), who found that 62% solved with the diagonal line condition. The most likely explanation for the difference is simply one of sampling: although all participants were naïve to the 9-dot problem, those in Experiment 4 were members of the general public who were recruited as unpaid volunteers during Open Days, while those in Experiment 5 were undergraduate students studying psychology, who were paid to take part. In terms of fit to the model, observed differences between conditions, the emergence of promising states and their predictiveness of eventual success were nearly identical to those in Experiment 4. The different participant populations would cause only the parameter k, and hence overall level of performance, to vary between experiments.

General Discussion

This paper has presented a model for predicting human performance on the nine-dot problem and a range of variants constructed by adding dots to and subtracting them from the standard nine-dot array. The model has two main components, a line maximization heuristic that selects moves on the basis of maximizing the number of dots cancelled in each move, and a progress monitoring heuristic that creates an impulse to seek alternative operators when a criterion for satisfactory progress is violated. Five experiments were conducted to test the following predictions deriving from the model:

(a) that performance is facilitated by the addition of extra dots that do not fundamentally change the shape of the required solution, but that encourage the early selection of moves lying on a correct solution path. The linear relationship found between solution rate and number of dots across the 11-, 12- and 13-dot problems in Experiment 1 supports this prediction, and appears to lie beyond the range of application of previous explanations of the nine-dot problem.

(b) that line maximization, rather than salience of starting point, determines performance in problems that do not require line extension to a non-dot point, a prediction confirmed by the results of Experiment 2.

(c) that making line extensions to non-dot points is a function of the impulse to seek emergent moves that itself results from the early violation of the criterion set by a progress-monitoring heuristic, a prediction confirmed by the results of Experiment 3;

(d) that violation of a progress-monitoring criterion, rather than the figural nature of the problem array, is the source of the impulse to seek emergent moves. This prediction was confirmed in Experiments 4 and 5, using standard nine-dot problems in which alternative first lines of a solution were given. Greater facilitation was found with a first line that remained within the implied figural boundary of the problem (that led to early criterion failure), relative to lines that extended beyond the implied boundary (that did not cause early criterion failure);

(e) that, once a criterion failure has invoked constraint relaxation allowing the inclusion in moves of non-dot points, the application of line maximization will yield 'promising states', that is, novel moves that can be capitalized upon as a source of insightful moves by the solver. This prediction was confirmed in Experiments 4 and 5, in which the generation of promising states served as a strong predictor of eventual success.

All these predictions discriminate between the model and the Gestalt explanation of figural integrity, the hypothesis-testing explanation of Weisberg and Alba (1981a), and the constraint-imposition explanation of Lung and Dominowski (1985). While the first three might be derived from a Gestalt account of performance, the fourth and fifth predictions are unique to the model. The present analysis also explains why solutions to the nine-dot problem are rare. With lookahead capacities of one, two and three lines, the motivation to seek alternative moves never arises until too late. This is because the best available moves always meet or surpass the current criterion up until the final move, by which time it is impossible to solve. Only in the case of someone looking four lines ahead does the best move fail to reach the criterion, and then only by a matter of one dot.

An important component of the way in which the model has been tested in this paper concerns the mix of lookahead types in the participant population. Experiments 3, 4 and 5 utilised estimates of lookahead estimated from the data of Experiments 1 and 2. While this approach has value, it is clear that the issue of lookahead will require further attention. One future approach may be to adopt alternative methods such as protocol analysis of move-by-move performance.

Ohlsson (1990) offers a conceptual framework for insight problem-solving based on the proposition that it appears in the context of an impasse that is unwarranted (in the sense that the person is competent to solve the problem). He raises three questions from this proposition: 1. How does the impasse arise? 2. How is the impasse breached? 3. What happens after the breach? Our model allows us to address these questions with greater precision than previous accounts. First, we propose that an impasse is reached, not simply when all the moves within a constrained problem space are exhausted (e.g., Issak & Just, 1995), but when there are no moves left that fit the current locally rational operator and that also generate satisfactory progress. Second, the impasse is breached when criterion failure signals the search for alternative operators. Third, our model suggests that the search for alternative operators is likely to be constrained by the re-application of the maximization heuristic. This gives rise to promising states, arising from unsuccessful attempts at applying alternative operators, that have value in meeting the progress-monitoring criterion and that are therefore retained in further attempts. In this way, the model provides an account of how insight may be achieved incrementally through experience of one or many partial solutions.

Promising states provide a further source of insightful moves that complements those specified by Knoblich, Ohlsson, Haider & Rhenius (1999) in their study of matchstick algebra problem-solving. Knoblich et al argue that two sources of insight are constraint relaxation (e.g., definitions of what constitute legitimate numerical operations) and chunk decomposition (e.g., chunks represented by sets of matchsticks organized as a common number pattern). Importantly, the order in which constraints are relaxed and chunks are decomposed is not random: minor constraints that affect few aspects of the current problem space are relaxed before key constraints, and loose chunks are decomposed before tight chunks. In the same way, promising states rarely emerge at random, and if they do, they are unlikely to be capitalized upon successfully in the absence of maximization and progress-monitoring. Instead, they are more likely to emerge, and to be recognized as being of value, when the problem space is attenuated, as it is by maximization and progress-monitoring.

The nine-dot problem is not the only example of an insight problem where the two general model principles, identifying locally rational operators, and monitoring progress against a criterion, are of value in explaining problem performance. For example, in the case of the six match problem (Weisberg & Alba, 1981a), a locally rational operator might be to form two sides of a triangle from each match, and a progress monitoring criterion might be the ratio of triangles completed to matches remaining. However, the principles are unlikely to apply equally to all problems that have been classed as insight problems. For example, the matchstick algebra problems used by Knoblich et al (1999) have only a single solution step. Thus, a progress monitoring heuristic is likely to be inappropriate.

The model provides an information processing account of insight in dot problems that is consistent with Weisberg's (1995) view that the nine-dot problem may be tackled through trial and error. However, trial and error solutions appear to be very rare. The model is also consistent with Ohlsson's (1990) framework for understanding insight in terms of dealing with impasses. An important extension that the model proposes is that the insightful move can be sparked by the anticipation of failure and not just by actual failure.

While the present model belongs in the information-processing tradition, it appears to us to represent a significant departure from previous outlines of possible models (Lung &

Dominowski, 1985; Wickelgren, 1974). The main difference is that previous attempts provided explanations of *failure to solve* the nine-dot problem, whereas the present approach explains not only the widespread failure to solve but also the occasional success. These occasional successes arise because, we propose, there are some circumstances where people may seek to apply operators (i.e. to seek novel moves) beyond those prescribed by the basic model. To allow for the search and application of new operators, a mechanism is proposed whereby people monitor progress on the task against a criterion. Should progress fail to meet the criterion then, it is proposed, an impulse arises to search for other operators, an impulse whose strength is proportional to the shortfall. The question then remains as to why some people may seek to apply operators beyond those prescribed by the basic model when solving the nine-dot problem in the absence of criterion failure. We suggest that an answer may lie in the parameterization of the model. As a simplifying assumption, we assumed that the parameter value, k, modifying the impulse to seek solutions, m, is constant in any particular problem variant. However individual differences in problem-solving experience, cognitive styles and so forth may cause k (and hence m) to vary.

Another possible mechanism for abandoning the current heuristic is through actual failure, rather than the anticipation of failure suggested here. While it seems likely that actual failure over repeated attempts would have such an effect (Kaplan & Simon, 1990), and is indeed likely to alter the parameter k, above, the proposal that the anticipation of failure also has this effect continues to have merit. First, it explains how successes with problems involving non-dot moves can arise on the first trial. Second, it is consistent with other evidence, that people may monitor the likely outcomes of current strategies (Larkin, 1983; VanLehn, 1989). Third, it explains the pattern of differences in performance between the “missing-dot” problems observed in the present Experiment 3, and the counter-intuitive facilitation demonstrated in Experiments 4 and 5.

The results of Weisberg & Alba (1981a, Experiment 1a), in which participants were given 100 attempts at the nine-dot but still failed to find a solution, suggest that the anticipation of failure may have qualitatively different outcomes from the experience of failure itself. In essence, knowing that current approaches have failed provides no guidance to the participant as to what kinds of constraint to relax, or what kinds of alternative operator to attempt. Moreover, alternative operators that are selected and tested at random in the absence of a progress-monitoring criterion will typically lead simply to further experience of failure, without offering any feedback as to the relative potential merits of each move. This may explain why simply instructing participants to "go outside the square" using the exhausted hint does not yield insight, and why participants who do try line extensions beyond the dots in the absence of criterion failure frequently return to executing moves that lie within the dot array. If participants capitalize upon the exhausted hint by abandoning operators that invoke line maximization in favor of other operators chosen at random, then they will also abandon the mechanisms by which they can judge the relative merits of novel moves. The application of maximization and concurrent progress-monitoring heuristics, and the subsequent controlled relaxation of problem constraints that occurs when the criterion is violated, effectively constrains the search for new problem spaces. More specifically, stages 1 and 2 of the model combine to constrain the participants at all times to a limited exploration of the problem space. When a criterion failure occurs, exploration may generate promising states.

In its adherence to the general principle of means-ends analysis, the model is less problem-specific than previous accounts of the processes underlying problem difficulty and the emergence of insight in the nine-dot problem. The generalization of means-ends analysis to a classic insight problem is made possible through the selection and application of locally-rational operators. In problems such as this, unlike well-defined problems like the Towers of Hanoi, progress toward a concrete and visualizable goal state cannot be monitored. Instead, progress monitoring must be criterion-based. One of the central features of the present model, the maximization heuristic, greatly attenuates the problem space, making an information-processing strategy feasible even for a processor with severely limited capacity. The proposition that people use such an heuristic is supported both theoretically and empirically. The present model may be viewed as filling a lacuna in the application of information-processing models to problem solving, by extending their range to the nine-dot problem. In some quarters the potential for doing so has long been anticipated (Wickelgren, 1974). In others, it has been considered unlikely to succeed, because of the potential complexity of the problem (Weisberg & Alba, 1981a).

References

Atwood, M. E. & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, *8*, 191-216.

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

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

Dominowski, R. L. (1981). Comment on "An examination of the alleged role of 'fixation' in the solution of several 'insight' problems" by Weisberg and Alba. *Journal of Experimental Psychology: General*, *110*, 193-198.

Ellen, P. (1982). Direction, past experience, and hints in creative problem solving: Reply to Weisberg and Alba. *Journal of Experimental Psychology: General*, *111*, 316-325.

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

Isaak, M. I. & Just, M. A. (1995). Constraints on thinking in insight and invention. In R. J. Sternberg and J. E. Davidson. *The nature of insight*. Cambridge, Mass.: MIT Press.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology*, *9*, 412- 440.

Kaplan, C. A. & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 373-419.

Karat, J. (1982). A model of problem solving with incomplete constraint knowledge. *Cognitive Psychology*, *14*, 538-559.

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),* 1534-1556.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner and A. Collins, eds. *Mental Models*. Hillsdale, NJ: Erlbaum.

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*, 804-811.

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

Ohlsson, S. (1990). Information-processing explanations of insight and related phenomena. In M. T. Keane and K. J. Gilhooly (Eds). *Advances in the psychology of thinking*, Vol. 1. London: Harvester Wheatsheaf.

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.

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem solving task. *Cognitive Psychology*, 8, 86-97.

Thomas, J. C. (1974). An analysis of behavior in the Hobbits-Orcs problem. *Cognitive Psychology*, 6, 257-269.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner, (Ed.) *Foundations of cognitive science*. Cambridge, Mass.: MIT Press.

Weisberg, R. W. (1995). Prolegomena to theories of insight problem solving: A taxonomy of problems. In R. J. Sternberg and J. E. Davidson. *The nature of insight.* Cambridge, Mass.: MIT Press.

Weisberg, R. W., & Alba, J. W. (1981a). An examination of the alleged role of “fixation” in the solution of several “insight” problems. *Journal of Experimental Psychology: General*, *110*,169-192.

Weisberg, R.W. & Alba, J.W. (1981b) Gestalt theory, insight and past experience: Reply to Dominowski. *Journal of Experimental Psychology: General, 110*,193-198.

Wertheimer, M. (1985). A Gestalt perspective on computer simulations of cognitive processes. *Computers in human behavior*, *1*, 19-33.

Wickelgren, W. (1974). *How to solve problems.* New York: Freeman.

Appendix

The Appendix provides details of the application of the model to each problem used in the experiments. The model's analysis is presented as a series of tables that follow the same general format.

The first row is labeled "Move Number". Between one and four moves are possible, depending on the value of lookahead and whether correct solutions can be found.

The second row is labeled "Maximum Value". Figures show the largest number of dots that can be cancelled by each move under Stage 1 of the model (i.e. assuming no line extensions to non-dot points).

The third row (Experiments 3, 4 & 5 only) is labeled "Criterion". Figures show the number of dots that must be cancelled by that move in order to meet the progress-monitoring criterion.

The fourth row (Experiments 3, 4 & 5 only) is labeled "Impulse Strength". Figures show the value of the impulse strength, m, operating on that move.

Subsequent rows show each of the available moves that have a maximum or criterion value. The numbers refer to dot positions as labeled in Figures 2-5. Where a move includes a line extension to a non-dot point, the number is shown in parentheses. Moves on a correct solution path are labeled "√". Moves on an incorrect solution path are labeled "x".

The final line in each table computes the probability of success after each move. Beneath each table is shown the joint probability of correct solution across moves.

Application of Model, Stage 1, to the stimuli of Experiment 1

13-Dot Problem,1 Lookahead

Move Number	1	2	3	4
Maximum Value	5	3	3	2
	13-1	13-1-10	13-1-10-11	13-1-10-11-1
				13-1-10-11-4
		13-1-11	13-1-11-10	13-1-11-10-1
				13-1-11-10-2
	1-13x			
probability	.5	1	1	1

joint probability correct = .5

2 Lookahead

Move Number	1	2
Maximum Value	8	5
	13-1-10	13-1-10-11-1
		13-1-10-11-4
	13-1-11	13-1-11-10-1
		13-1-11-10-2
	10-1-13x	
	11-1-13x	
probability	.5	1

joint probability correct = .5

3 Lookahead

Move Number	1	2
Maximum Value	11	2
	13-1-10-11	13-1-10-11-1
		13-1-10-11-4
	13-1-11-10	13-1-11-10-1
		13-1-11-10-2
	10-11-1-13x	
	11-10-1-13x	
probability	.5	1

joint probability correct = .5

4 Lookahead

Move Number	1
Maximum Value	13
	13-1-10-11-1
	13-1-10-11-4
	13-1-11-10-1
	13-1-11-10-2
	1-10-11-1-13
	1-11-10-1-13
probability	1

probability correct = 1

12-Dot Problem, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	4	3	3 or 2	2 or 3
	12-1	12-1-10	12-1-10-11	12-1-10-11-1
				12-1-10-11-4
		12-1-11	12-1-11-10	12-1-11-10-1
				12-1-11-10-2
	1-10	1-10-11	1-10-11-1	1-10-11-1-12
			1-10-11-4x	
	1-11	1-11-10	1-11-10-1	1-11-10-1-12
			1-11-10-2x	
	1-12x			
	10-1x			
	11-1x			
	10-11x			
	11-10x			
probability	.375	1	.667	1

joint probability correct = .25

2 Lookahead

Move Number	1	2
Maximum Value	7	5
	12-1-10	12-1-10-11-1
		12-1-10-11-4
	12-1-11	12-1-11-10-1
		12-1-11-10-2
	1-10-11	1-10-11-1-12
	1-11-10	1-11-10-1-12
	10-1-11x	
	11-1-10x	
	10-11-1x	
	11-10-1x	
	10-1-12x	
	11-1-12x	
probability	.4	1

joint probability correct = .4

<u>3 Lookahead</u>

Move Number	1	2
Maximum Value	10	2
	12-1-10-11	12-1-10-11-1
		12-1-10-11-4
	12-1-11-10	12-1-11-10-1
		12-1-11-10-2
	10-11-1-12x	
	11-10-1-12x	
probability	.5	1

joint probability correct = .5

<u>4 Lookahead</u>

Move Number	1
Maximum Value	12
	12-1-10-11-1
	12-1-10-11-4
	12-1-11-10-1
	12-1-11-10-2
	1-10-11-1-12
	1-11-10-1-12
probability	1

joint probability correct = 1

<u>11-dot problem, 1 Lookahead</u>

Move Number	1	2	3	4
Maximum Value	4	3	2	2
	1-10	1-10-11	1-10-11-1	1-10-11-1-9
			1-10-11-4x	
	1-11	1-11-10	1-11-10-1	1-11-10-1-9
			1-11-10-2x	
	10-1x			
	11-1x			
	10-11x			
	11-10x			
probability	.333	1	.5	1

joint probability correct=.17

2 Lookahead

Move Number	1	2
Maximum	7	4
	1-10-11	1-10-11-1-9
	1-11-10	1-11-10-1-9
	10-1-11x	
	11-1-10x	
	10-11-1x	
	11-10-1x	
probability	.333	1

joint probability correct=.33

3 Lookahead

Move number	1	2
Maximum	9	2
	1-11-10-1	1 9
	1-10-11-1	1-9
	9-1-11-10	10-1
		10-2
	9-1-10-11	11-1
		11-4
	1-11-10-2x	
	1-10-11-4x	
	10-11-1-10x	
	10-11-1-3x	
	11-10-1-11x	
	11-10-1-7x	
	10-1-11-10x	
	10-1-11-6x	
	11-1-10-11x	
	11-1-10-8x	
probability	.29	1

joint probability correct =. 29

4 Lookahead

Move Number	1
Maximum Value	11
	9-1-10-11-1
	9-1-10-11-4
	9-1-11-10-1
	9-1-11-10-2
	1-10-11-1-9
	1-11-10-1-9
probability	1

probability correct=1

Application of Model, Stage 1, to the stimuli of Experiment 2

Problem 13 A, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	6	3	2	2
	13-11	13-11-10	13-11-10-1	13-11-10-1-9
			13-11-10-2x	13-1-10-11-4
	11-13x			
probability	.5	1	.5	1

joint probability correct=.25

2 Lookahead

Move Number	1	2
Maximum Value	9	4
	13-11-10	13-11-10-1-9
	10-11-13x	
probability	.5	1

joint probability correct = .5

3 Lookahead

Move Number	1	2
Maximum Value	11	2
	13-11-10-1	13-11-10-1-
	13-11-10-2x	
	1-10-11-13x	
probability	.33	1

joint probability correct = .33

4 Lookahead

Move Number	1
Maximum Value	13
	13-11-10-1-9
	9-1-10-11-13
probability	1

probability correct = 1

Problem 13B, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	5	3	2	2
	13-11	13-11-10	13-11-10-1	13-11-10-1-12
			13-11-10-2x	
	11-13x			
probability	.5	1	.5	1

joint probability correct = .25

2 Lookahead

Move Number	1	2
Maximum Value	8	5
	13-11-10	13-11-10-1-12
	10-11-13x	
probability	.5	1

joint probability correct = .5

3 Lookahead

Move Number	1	2
Maximum Value	10	3
	13-11-10-1	13-11-10-1-12
	13-11-10-2x	
	1-10-11-13x	
probability	.33	1

joint probability correct = .33

4 Lookahead

Move Number	1
Maximum Value	13
	13-11-10-1-12
	12-1-10-11-13
probability	1

probability correct = 1

Application of Model, Stage 2, to the stimuli of Experiment 3

Problem 12A, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	4			
Criterion	3			
Impulse strength	0			
	13-5x			
	5-13x			
	10-11x			
	11-10x			
probability	0			

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	6	3
Criterion	6	6
Impulse strength	0	.5k
	2-10-11	2-10-11-(1)-13
	4-11-10	4-11-10-(1)-13
	11-10-2x	
	10-11-4x	
probability	.5	1x.5k

joint probability correct=.25k

3 Lookahead

Move Number	1	2
Maximum Value	8	2
Criterion	9	2
Impulse strength	.11k	0
	13-(1)-11-10	13-(1)-11-10-(1)
		13-(1)-11-10-2
	13-(1)-10-11	13-(1)-10-11-(1)
		13-(1)-10-11-4
	10-11-(1)-13x	
	11-10-(1)-13x	
probability	.5x.11k	1

joint probability correct = .06k

4 Lookahead

Move Number	1
Maximum Value	10
Criterion	12
Impulse strength	.17k
	13-(1)-10-11-(1)
	13-(1)-10-11-4
	13-(1)-11-10-(1)
	13-(1)-11-10-2
	4-11-10-(1)-13
	2-10-11-(1)-13
probability	1x.17k

probability correct =.17k

Problem 12B, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	5	3	2	
Criterion	3	2.33	2	
Impulse strength	0	0	0	
	13-11	13-11-10	13-11-10-2x	
	11-13x			
probability	.5	1	0	

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	8	3
Criterion	6	4
Impulse strength	0	.25k
	13-11-10	13-11-10-(1)-9
	10-11-13x	
probability	.5	1x.25k

joint probability correct = .125k

3 Lookahead

Move Number	1	2
Maximum Value	10	
Criterion	9	
Impulse strength	0	
	13-11-10-2x	
	2-10-11-13x	
probability	0	

joint probability correct = 0

4 Lookahead

Move Number	1
Maximum Value	11
Criterion	12
Impulse strength	.08k
	13-11-10-(1)-9
	9-(1)-10-11-13
probability	1x.08k

probability correct=.08k

Problem 12C, 1 Lookahead

Move Number	1	2	3	4
Maximum Value	5	3	2	
Criterion	3	2.33	2	
Impulse strength	0	0	0	
	13-1	13-1-11	13-1-11-6x	
	1-13x			
probability	.5	1	0	

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	8	3
Criterion	6	4
Impulse strength	0	.25k
	13-1-11	13-1-11-(10)-1
		13-1-11-(10)-2
	11-1-13x	
probability	.5	1x.25k

joint probability correct=.13k

3 Lookahead

Move Number	1	2
Maximum Value	10	
Criterion	9	
Impulse strength	0	
	13-1-11-6x	
	6-11-1-13x	
probability	0	

joint probability correct = 0

4 Lookahead

Move Number	1
Maximum Value	11
Criterion	12
Impulse strength	.08k
	13-1-11-(10)-1
	13-1-11-(10)-2
	1-11-(10)-1-13
	1-(10)-11-1-13
	13-1-(10)-11-1
	13-1-(10)-11-4
probability	1x.08k

probability correct = .08k

Application of Model Stage 2 to the stimuli of Experiments 4 and 5.

Problem 9A, 1 Lookahead

Move Number	1	2	3
Maximum Value	2		
Criterion	2		
Impulse strength	0		
	1-7x		
	1-9x		
	10-8x		
probability	0.0		

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	4	
Criterion	4	
Impulse strength	0	
	1-7-9x	
	1-9-7x	
probability	0.0	

joint probability correct=0

3 Lookahead

Move Number	1
Maximum Value	5
Criterion	6
Impulse strength	.17k
	10-(11)-1-9
probability	1x.17k

joint probability correct = .17k

Problem 9B, 1 Lookahead

Move Number	1	2	3
Maximum Value	2		
Criterion	2		
Impulse strength	0		
	1-7x		
	1-3x		
probability	0.0		

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	3	
Criterion	4	
Impulse strength	.25k	
	1-(10)-8x	
	1-(11)-6x	
probability	0.0	

joint probability correct=0

3 Lookahead

Move Number	1
Maximum Value	5
Criterion	6
Impulse strength	.17k
	1-(10)-(11)-4
	1-(11)-(10)-1
	1-(11)-(10)- 2
	1-(10)-(11)-1
probability	1x .17k

joint probability correct = .17k

Problem 9C, 1 Lookahead

Move Number	1	2	3
Maximum Value	2		
Criterion	2		
Impulse strength	0		
	1-7x		
	1-3x		
	9-7x		
	9-3x		
probability	0.0		

joint probability correct = 0

2 Lookahead

Move Number	1	2
Maximum Value	3	
Criterion	4	
Impulse strength	.25k	
	1-(10)-8x	
	1-(11)-6x	
	9-(12)-4x	
	9-(13)-2x	
probability	0.0	

joint probability correct=0

3 Lookahead

Move Number	1
Maximum Value	5
Criterion	6
Impulse strength	.17k
	1-(10)-(11)-4
	1-(10)-(11)-1
	1-(11)-(10)- 2
	1-(11)-(10)-1
	9-(12)-(13)-8
	9-(12)-(13)-9
	9-(13)-(12)-6
	9-(13)-(12)-9
probability	1x .17k

joint probability correct = .17k

Tables

Table 1: The model's performance in the four problems of Experiment 1

Problem	Dependent variable	Lookahead 1	2	3	4
9-dot	probability of success	.00	.00	.00	.11k
	mean trials to succeed	-	-	-	≥9.1
	percentage solving in 10 trials	0	0	0	0-69
11-dot	probability of success	.17	.33	.29	1.00
	mean trials to succeed	6	3	3.5	1
	percentage solving in 10 trials	84	98	97	100
12-dot	probability of success	.25	.40	.50	1.00
	mean trials to succeed	4	2.50	2	1
	percentage solving in 10 trials	94	99	100	100
13-dot	probability of success	.50	.50	.50	1.00
	mean trials to succeed	2	2	2	1
	percentage solving in 10 trials	100	100	100	100

Table 2: The number of participants in each condition, the percentage solving on the first attempt, the percentage solving in 10 attempts and the mean number of trials to solve, Experiment 1.

Problem	N	Percent solving in one trial	Percent solving in 10 trials	Mean number of trials to solve
9-dot	27	0	0	-
11-dot	30	50	93	1.86
12-dot	25	60	80	1.40
13-dot	30	73	90	1.22

Table 3 The percentage of participants starting first solution attempts from different locations (and the percentage solving), Experiment 1.

	Problem			
Location	13-dot	12-dot	11-dot	9-dot
Bottom right	68 (61)	55 (55)	58 (54)	0 (-)
Top left	25 (14)	30 (15)	13 (8)	43 (0)
Bottom left	4 (0)	5 (0)	17 (0)	43 (0)
Top right	4 (0)	5 (0)	8 (0)	5 (0)
Other	0 (-)	5 (0)	4 (0)	10 (0)
Number reporting starting point	28	20	24	21

Table 4: Percentage of participants starting from different locations, (and percentage solving), Experiment 2.

Start location	Problem 13	13A	13B
Bottom right	74 (74)	13 (13)	5 (5)
Top left	17 (17)	74 (48)	75 (65)
Bottom left	0 (-)	9 (0)	0 (-)
Top right	4 (0)	4 (0)	15 (-)
Other	4 (0)	0 (-)	5 (0)
Number reporting starting point	23	23	20

Table 5: Percentage of participants starting from different locations, Experiment 3. N= number of participants in that condition reporting a starting point.

Problem					
12A		12B		12C	
Start	% (N=43)	Start	% (N=53)	Start	% (N=43)
13	60	13	62	13	53
10	12	11	15	11	30
11	9	10	15	1	12
2	12	2	2		
4	2	7	2		
		9	2		
Other	5	Other	2	Other	5

Figure Captions

Figure 1: The nine-dot problem and its solution

Figure 2: The 9-,11-, 12- and 13-dot problems from Experiment 1, as seen by participants (upper panel) and in numerical notation (lower panel)

Figure 3: Problems 13, 13A and 13B from Experiment 2, as seen by participants (upper panel) and in numerical notation (lower panel)

Figure 4: Problems 12A, 12B and 12 C from Experiment 3, as seen by participants (upper panel) and in numerical notation (lower panel)

Figure 5: Problems 9A, 9B and 9C from Experiments 4 and 5, as seen by participants (upper panel) and in numerical notation (lower panel)

Figure 6. An illustration of a "promising state" in Experiments 4 and 5. The dotted line represents the line given in condition 9C. The solid lines represent a two-lookahead move from dot 1, through dot 7, turning on a non-dot point, then through dot 8 to dot 6. This move meets the two-lookahead criterion (cancel four dots) but does not in itself guarantee a solution to the problem.

Figure 1:

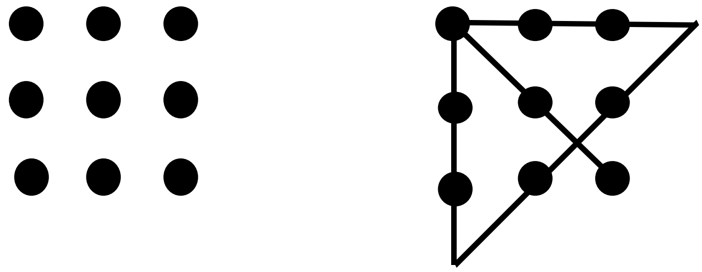


Figure 2:

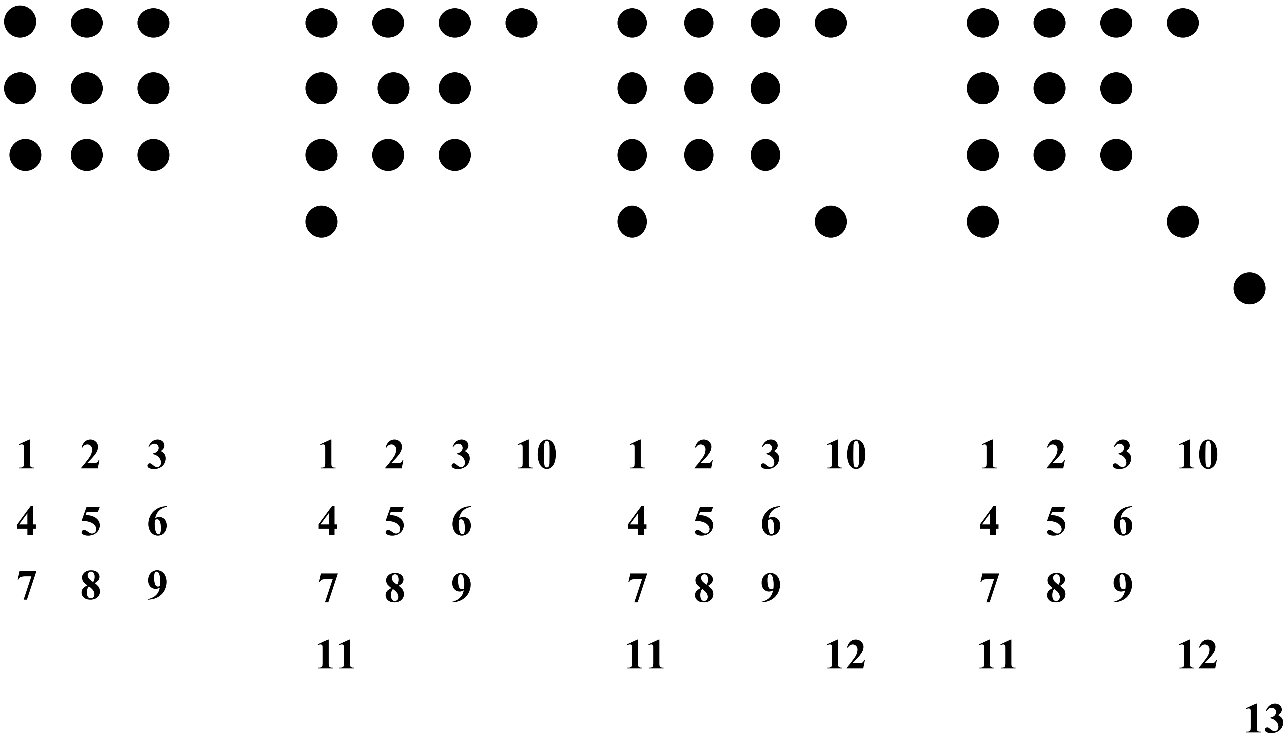

1 2 3
4 5 6
7 8 9
1 2 3 10
4 5 6
7 8 9
11
1 2 3 10
4 5 6
7 8 9
11 12
1 2 3 10
4 5 6
7 8 9
11 12
13

Figure 3:

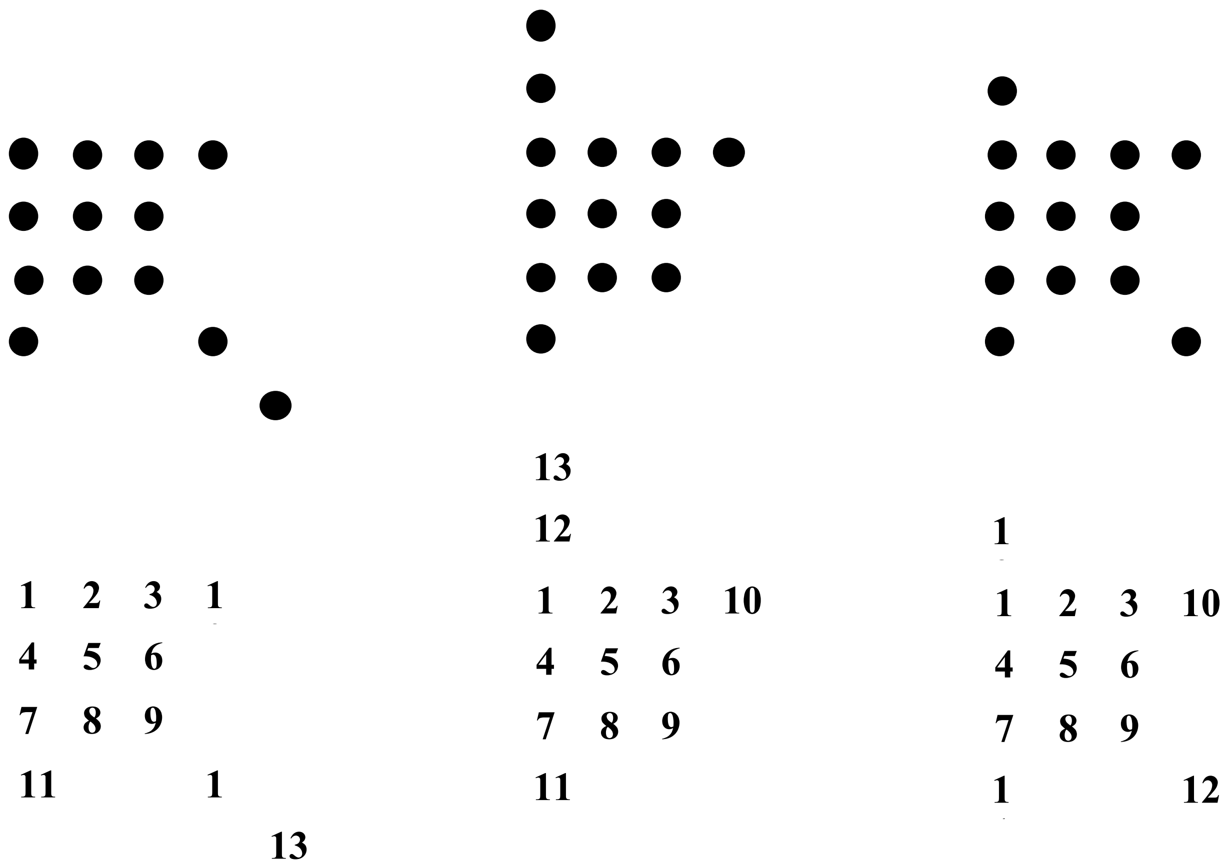


Figure 4:

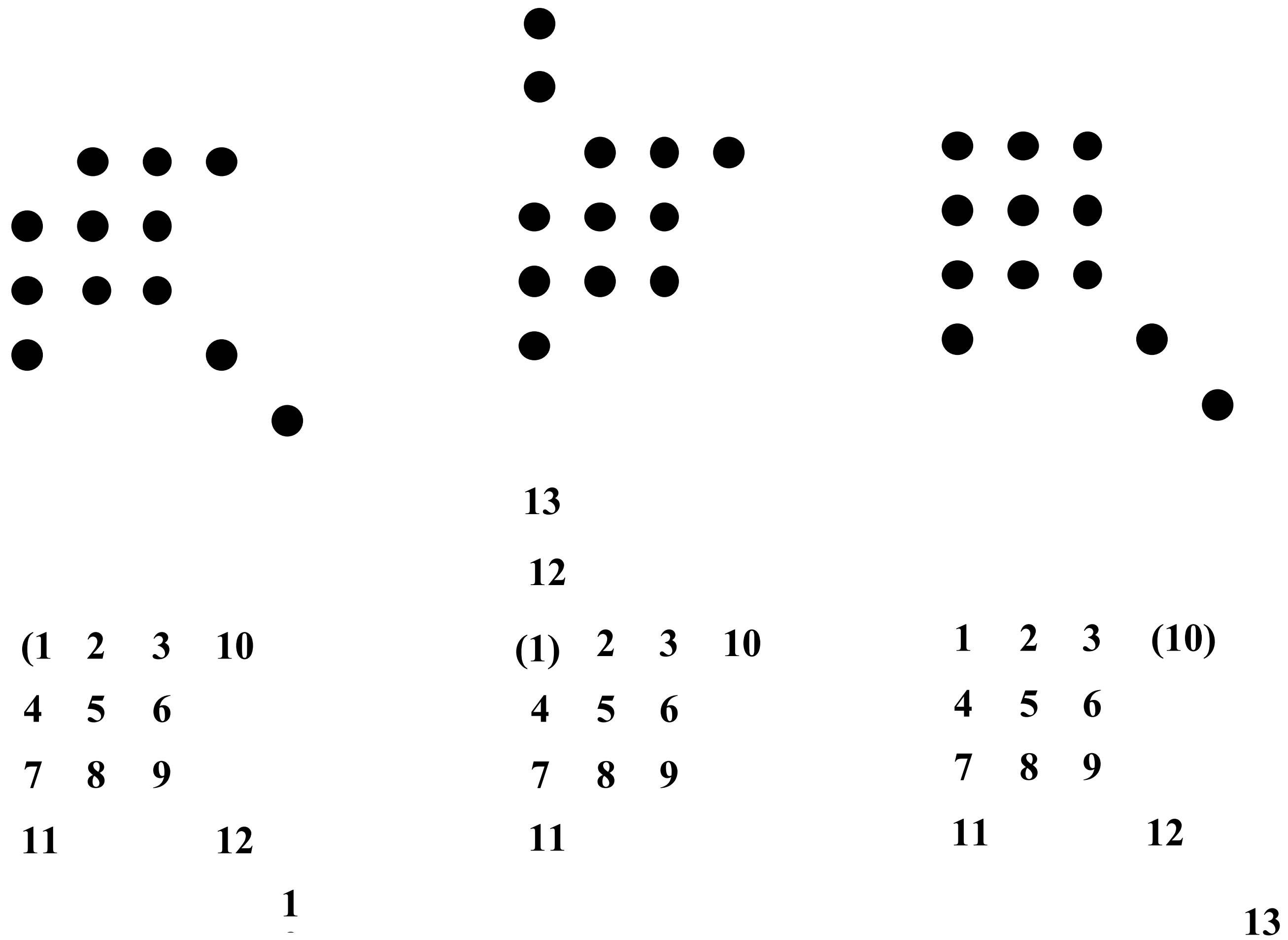

13
12
(1 2 3 10
4 5 6
7 8 9
11 12
1
(1) 2 3 10
4 5 6
7 8 9
11
1 2 3 (10)
4 5 6
7 8 9
11 12
13

Figure 5:

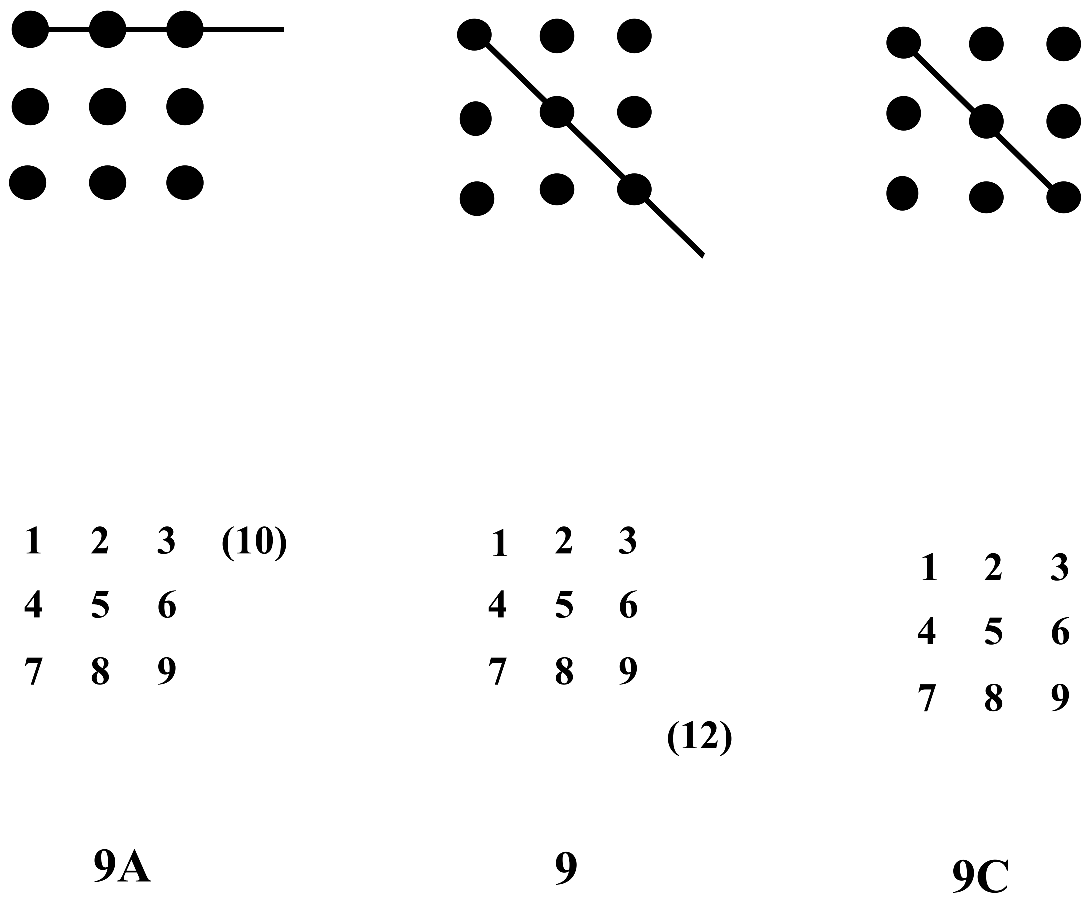
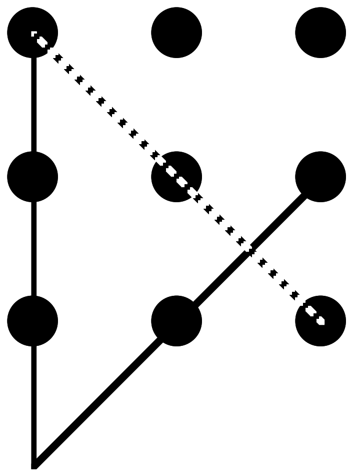


Figure 6:



1	2	3
4	5	6
7	8	9