An eye movement study of insight problem solving

GÜNTHER KNOBLICH
Max Planck Institute for Psychological Research, Munich, Germany

and

STELLAN OHLSSON and GARY E. RANEY
University of Illinois, Chicago, Illinois

The representational change theory of insight claims that insight problems cause impasses because they mislead problem solvers into constructing inappropriate initial representations. Insight is attained when the initial representation is changed. In the present study ($N = 24$), we tested three specific implications of these hypotheses against eye movements recorded while participants solved matchstick arithmetic problems. The results were consistent with the predictions, providing converging evidence with prior findings using solution rates and solution times. Alternative theories of insight can explain individual findings, but only the representational change theory accounts for both the performance data and the eye movement data. The present study also suggests that eye movement recordings provide an important new window into processes of insight problem solving.

Although unfamiliar problems can sometimes be solved by working through them step by step at a steady pace (e.g., Anderson & Lebiere, 1998; Newell & Simon, 1972), solutions to insight problems exhibit a characteristic temporal pattern: An initial period of purposeful problem solving activity is followed by an impasse, a state of mind in which the problem solver feels that all options have been explored and he or she cannot think of what to do next. Continued concentration on the problem will, in a proportion of cases, cause a new idea or option to come to mind. This so-called aha-experience is typically unanticipated by the problem solver (Metcalfe, 1986) and is followed by rapid completion of the solution, at least for the simple problems that are used to study insight in the laboratory. Well-known examples of insight problems include the *six matches problem* (Katona, 1940), the *two-string problem* (Maier, 1931), the *candle problem* (Weisberg & Suls, 1973), and the *nine-dot problem* (MacGregor, Ormerod, & Chronicle, 2001; Scheerer, 1963). After a pause in the decades after the cognitive revolution in the 1950s, there has been a recent resurgence of interest in insight (e.g., Smith, Ward, & Finke, 1995; Sternberg & Davidson, 1995).

The impasse–insight sequence poses a double theoretical challenge (Ohlsson, 1992). The first half of that challenge is to understand why people become stuck on problems for which they have the necessary knowledge. Many insight problems have very short solution paths and require only concepts and skills that are well within the competence of the average educated adult. Nevertheless, the impasses produced by such problems can last for several minutes or even hours. If a person is competent to solve a problem, why does he or she experience an impasse? The second half of the theoretical challenge is to explain how impasses are resolved. If there is some cognitive mechanism or factor that prevents a person from applying his or her knowledge to a problem, why is the impasse not permanent? What are the cognitive processes that allow a person to break out of an impasse?

Theories of insight problem solving tend to emphasize one or the other of these two challenges. The *functional fixedness* hypothesis claims that mental representations of objects are associated with the common functions for these objects. If in a certain problem, a familiar object has to be used in an unfamiliar way, the spontaneous retrieval of its familiar function blocks its unfamiliar usage and an impasse arises (Duncker, 1945; Keane, 1989). The *mental ruts* hypothesis (Smith, 1995) claims that impasses are encountered because the repeated exploration of an unsuccessful search path or the search for the same knowledge element adds more and more activation to this path. This, in turn, decreases the probability of exploring another path and the problem solver gets stuck (see also Simon, 1966). This concept is very similar to the *Einstellung* concept introduced by Luchins and Luchins (1959). According to this

hypothesis, once a solution path has become habitual, the problem solver no longer searches the problem space for alternative and potentially more efficient solution paths. If the habitual path is blocked in some way, an impasse results. MacGregor et al. (2001) have recently proposed a principle of satisfactory progress that says that a person is experiencing an impasse on the nine-dot problem when he or she cannot find a next line to draw that picks up a sufficient number of dots. All four of these hypotheses explain why problem solvers enter impasses, but they do not explain how impasses are resolved.

According to Gestalt theory (Köhlich, 1924, 1925; Wertheimer, 1959), thinking begins when the perceptual field is in a state of imbalance or tension. A solution appears in consciousness when the perceptual field reorganizes itself into a better, more harmonious or balanced state (Ohlsson, 1984a). Another type of explanation is the Darwinian process of variation and selection (Simonton, 1988, 1995). The variation–selection hypothesis claims that novel solutions are constructed by creating more or less random variants of existing solutions and then evaluating those variants until one is found that solves the current problem. Others have described insight problem solving as a gradual transformation of past experience (Perkins, 1981; Weisberg, 1986; Weisberg & Alba, 1981). According to this view, creative products or solutions result from gradual transformations of past experience. These three hypotheses explain how people can succeed in solving unfamiliar problems. However, they do not explain why there are impasses. There is no clear reason why the reorganization of the perceptual field, the variation–selection process, or the gradual transformation of experience would come to a standstill or be delayed longer on some problems than on others.

We have developed a theory of insight that addresses both theoretical challenges (Knoblich, Ohlsson, Haider, & Rhenius, 1999; Ohlsson, 1984b, 1992). Our explanation for why people encounter impasses is that insight problems have a high probability of triggering an initial mental representation that has a low probability of leading to the solution. The problem as initially perceived by the problem solver interacts with his or her prior knowledge in such a way as to activate knowledge elements (i.e., analogies, concepts, ideas, operators, principles, rules, schemas, skills, etc.) that are not sufficient for solving it. It may also lead to inhibition or suppression of knowledge elements that are essential for the solution. The activation of unhelpful knowledge elements and the consequent inhibition of more helpful ones is the cause of impasses.

The second principle of our theory says that impasses are resolved by revising the unhelpful initial representation. The consequence of such a change is that the distribution of activation over memory is altered and the problem solver might access knowledge elements that have so far remained inactive but are essential for the solution. The appearance of a crucial but previously unheeded knowledge element in working memory (and hence in subjective experience) is the aha-experience. Problem representa-tions can be revised in multiple ways (Ohlsson, 1992), for instance by *constraint relaxation* (i.e., the deactivation of some knowledge element that has acted as a constraint on the options initially considered) and *chunk decomposition* (i.e., the separation of the components of a perceptual chunk; Knoblich et al., 1999).

Kaplan and Simon (1990) have proposed an alternative representational change theory of insight. They suggest that people encounter impasses because their initial problem representation is incomplete rather than misleading. Hence, the problem solver has to resolve an impasse by noticing and encoding additional features of the problem rather than by abandoning or deactivating unhelpful ones. In particular, they hypothesize that people search for and notice *invariants* (i.e., properties that remain the same when the problem is represented in different ways). Although our notion of activation redistribution claims that restructuring processes occur outside consciousness (Bowden & Beeman, 1998; Metcalfe, 1986), Kaplan and Simon (1990) suggest that those processes are under conscious control.

In the past, researchers have typically tested their insight theories against performance measures—primarily the time it takes to find a solution and the probability of finding a solution within a given time frame. Our own past work is no exception (Knoblich et al., 1999). However, as argued by Newell and Simon (1972) and others, performance measures provide only weak information about underlying processes and need to be supplemented with process tracing methods. Verbal protocols have been used to study insight (Kaplan & Simon, 1990; Schooler & Melcher, 1995), but Schooler, Ohlsson, and Brooks (1993) have questioned whether the think-aloud method is nonreactive vis a vis insight problem solving. Eye movements present a potentially powerful alternative. There is a close connection between which display item a person is looking at and which item he or she is thinking about, as well as between fixation duration and amount of processing (Just & Carpenter, 1976; Knoblich & Rhenius, 1995; Rayner, 1978, 1998). Eye movement data thus provides a temporally fine-grained record of problem solving processes. In spite of the proven usefulness of eye movement recordings, there are no studies that apply this methodology to insight problem solving. In the present study, we tested three predictions from our insight theory against eye movement data recorded while the participants solved match stick arithmetic problems.

Eye Movements in Match Stick Arithmetic

We used matchstick arithmetic problems, because each problem consists of a small number of distinct elements, arranged in a horizontal sequence. Hence, we could determine with precision which problem component a participant was fixating at any one moment in time. In matchstick arithmetic, the problem solver is faced with an incorrect arithmetic statement expressed in Roman numerals constructed out of matchsticks. The goal is to correct the arithmetic statement by moving a single matchstick from

one position in the statement to another. Three examples are shown in Figure 1. Problem A is solved by moving the vertical stick on the far left to the right hand side of the Roman numeral V, thereby creating the correct arithmetic statement, VI = III + III. Because all problems used in our study are solved with a single, physically trivial move, their degree of difficulty is solely a function of the probability of thinking of the right move. We invite the reader to solve Problems B and C in Figure 1 before reading further.

Unlike Problem A, Problem B cannot be solved by changing the values in the equation. Instead, one has to change the plus sign into a second equal sign to obtain the solution, III = III = III. In prior research, we have shown that Problem B is considerably more difficult than Problem A (Knoblich et al., 1999; Knoblich & Wartenberg, 1998). Our explanation is that matchstick arithmetic problems activate the problem solver's prior knowledge of arithmetic and that this knowledge biases the initial representation in such a way that values are encoded as variable elements but operators (i.e., plus, minus, and equal signs) are encoded as constants. In arithmetic, there are many operations that change the values in an equation, but few that change the operators. In fact, arbitrarily altering the operators in an equation usually violates the very meaning of that equation. Unless the constraint against altering the operators is relaxed, Problem B cannot be solved.

Like Problem A, Problem C is solved by changing a value. The solution is to slide the left-slanted stick that is part of the symbol X to the left, to obtain the solution VI = III + III. Roman numerals are meaningful chunks for our experimental participants. Unless the relevant chunks, IV and X, are decomposed into their components, Problems A and C cannot be solved. We have shown that Problem C is considerably more difficult to solve than Problem A (Knoblich et al., 1999). Our explanation is that it is more difficult to intellectually detach a component that has no meaning of its own, like the left-slanted stick in X, than a stick that carries its own meaning, like the "I" in IV.

If our account of insight problem solving in general and matchstick arithmetic solutions in particular is correct, what regularities can we expect in the participants' eye movements? We focus on three predictions.

First, impasses should affect the mean fixation duration. During an impasse, the problem solver is not actively exploring the problem space. He or she does not know what to do next, so he or she tends to stare at the problem without testing particular solution ideas. The result should be fewer eye movements per unit of time. Within a fixed time frame, this translates into longer fixations. Because we expect impasses in Problems B and C but not in Problem A, mean fixation duration should be greater in the first two problems. Furthermore, because impasses only occur after the initial exploration of the problem space, mean fixation duration should increase throughout the problem solving effort for Problems B and C, as more and more problem solvers enter impasses, but not for Problem A. Finally, this pattern should hold for both successful and unsuccessful problem solvers, because both are hypothesized to experience impasses. We have no reason to expect impasse-driven fixations to be aimed at one problem element rather than at another, so this is strictly a prediction about the duration of fixations over time, not a prediction about the allocation of attention.

Second, the initial representation of the problem should affect attention allocation. If the participants initially regard the values in the equations as the variable elements and the operators as constants, it follows that they should focus on the values during the initial, exploratory phase of problem solving. Hence, we predict that a larger proportion of the total fixation time should fall on the values (results and operands) than on the operators. This prediction holds only for the initial phase of problem solving, before the constraint against altering other problem elements is relaxed.
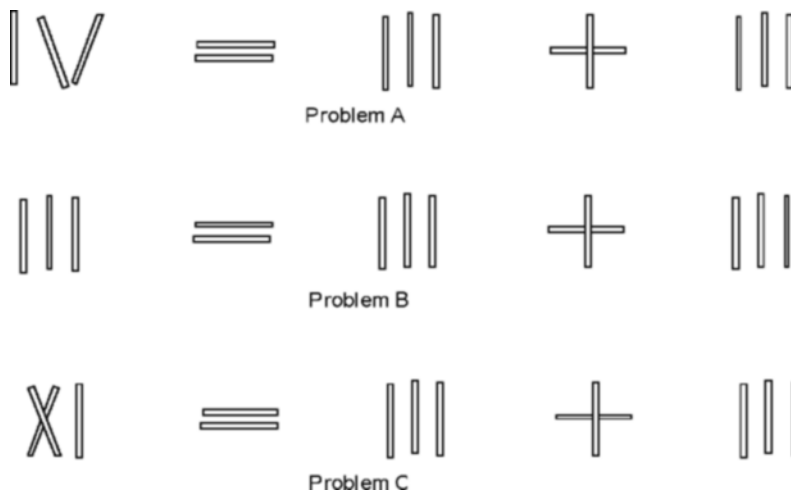

Figure 1. Matchstick arithmetic problems.

A recent study by Ballard, Hayhoe, Pook, and Rao (1997) suggests a functional distinction between short and long fixations. A short fixation primarily serves to reencode a problem element into working memory; it does not necessarily imply that the reencoded element is processed further. A longer fixation is more likely to signal deeper processing. To take this distinction into account, we make separate predictions for short and long fixations.

For short fixations, no differences are expected between different problem elements, because there is no reason why any single element in a matchstick arithmetic problem should need to be reencoded more often than another. Because there are five elements in each problem (result, equal sign, operand1, plus sign, operand2), we expect that, for short fixations, 20% of the fixation time will be spent on each problem element. There should be no differences between problems in this respect.

For long fixations, the percentage of fixation time spent on the values in the expression should exceed the chance level of 20%, and the percentage of fixation time spent on the operators should fall below that level. Again, there should be no differences between different problems, because the initial bias should be present whether it helps or hinders the problem solving process. To be clear, we are not predicting that fixations of values are longer, on average, than fixations of operators, but that a larger proportion of the total amount of time allotted to long fixations is spent looking at the values. Thus, this is a prediction about the allocation of attention, not about the duration of fixations.

Of course, focusing on the longer fixations poses the potential problem that the population of relevant fixations includes not only those associated with deep processing, but some that are associated with impasses as well. Clearly, meaningful shifts in attention can only be expected to affect the former and not the latter. A preponderance of impasse-driven long fixations will dilute the predicted effects and make them more difficult to observe in the data. For this reason, the prediction about the distribution of fixation time holds only for the initial, preimpasse phase of problem solving.

Third, the revision of the initial problem representation that precedes insight should also affect attention allocation. We make separate predictions for Problems B and C. In Problem B, the initial focus on the values implies that the percentage of fixation time spent on the arithmetic operators (the plus and equal signs) should be lower than the percentage spent on the values. This holds for both successful and unsuccessful problem solvers. For successful problem solvers, this bias should become less pronounced or should reverse when the constraint against altering the operators is relaxed. The implied increase in the percentage of fixation time spent on the operators should only happen for successful problem solvers. Unsuccessful problem solvers presumably fail because they never relax that constraint, so there is no reason to expect a reversal in attention allocation for them. The change in attention allocation should affect long fixations but not necessarily short ones.

In Problem C, the result is the crucial problem element. Initially, the percentage of fixation time spent on the result should be comparable to the percentage fixation time spent on other problem elements, and it should be the same for successful and unsuccessful problem solvers. When the crucial chunk X is decomposed, the percentage of fixation time spent on the result should increase. This effect should only occur in successful problem solvers. Unsuccessful problem solvers presumably fail because they never decompose that chunk, so there is no reason to expect a change in the time spent on the result for them. Again, the change in attention allocation should affect long fixations but not necessarily short ones.

Considered one by one, these predictions are not unique to our theory. On the one hand, the principles of functional fixedness, mental ruts, and Einstellung all predict impasses and hence the elongation of fixation times after the exploratory phase. However, because they do not predict changes in problem representation, they do not generate specific predictions about changes in attention allocation. On the other hand, principles like variation and selection and gradual transformation of past experience do not predict impasses. In addition, they are very general and lack any conceptual mechanism for relating them to the details of either tasks or behaviors, so it is difficult to derive precise predictions about attention allocation from them. The conjunction of our three predictions depends essentially on our analysis of exactly why people encounter impasses in matchstick arithmetic and exactly how their problem representations change when they resolve those impasses.

METHOD

Participants

Twenty-four undergraduates from the University of Illinois at Chicago, 10 of them male, participated in the study for course credit. They ranged in age from 18 to 29 years. We selected only participants who were familiar with Roman numerals. The participants were seen individually. Because of restrictions presented by the eye movement monitor, the participants could not wear glasses during the experiment, but were permitted to wear contact lenses. The participants reported no eye abnormalities (e.g., lazy eyes).

Apparatus

A Dr. Bouis Monocular Oculometer was used to track eye movements. Eye position was sampled at a rate of 1000 Hz. The initiation of a fixation was defined as the point when five consecutive samples each differed from the sample taken 5 msec earlier by less than 1/9 of a visual degree. The initiation of a saccade was defined as the point when three consecutive samples each differed from the previous sample by more than 1/9. Problems were displayed on a VGA monitor positioned 72 cm from the participant. The visual angle subtended by the problems was 12° horizontal and 2° vertical.

Procedure, Materials, and Design

The participants first went through a training phase to speed up the recognition of Roman numerals. Then, the participants were given a description of the eye tracking apparatus. Afterwards, the experimenter prepared a bite bar that served to increase the precision of the measurement by stabilizing the participants' heads. Finally, the participants received instructions about the rules of matchstick arithmetic and the experimental procedure. Before presenting

the first problem, a calibration procedure was conducted to adjust the eye tracker.

The participants solved three matchstick problems (Problems A, B, and C) that were presented in random order. They pushed a button as soon as they thought that they had found the solution to a problem, removed their teeth from the bite bar, and said the solution out loud. If the solution was correct, the next problem was presented. If not, they continued to work on the same problem until they found another solution or until they reached the time limit of 5 min. If the participants did not solve a problem within 5 min, they were told the solution.

A short recalibration was done (1) before presenting a new task, (2) after an incorrect solution had been proposed, and (3) when a participant had worked on a problem continuously for 1 min without proposing a solution. The recalibration took between 3 and 10 sec.

RESULTS

The dependent variables were the frequency of solution during a 5-min interval, the solution time for problems solved, and the position and the duration of each fixation on the display.

Solution Frequencies and Solution Times

The performance results confirmed the predicted differences in task difficulty. Figure 2 shows the cumulative frequency of solutions across 1-min intervals for Problems A, B, and C.

As Figure 2 shows, the differences between problems hold at each interval. Single chi-square tests were computed to test whether frequency of solution within a 5-min time limit differed significantly between two problems. Problem B, which requires that the operator be changed, was solved significantly less often (9 out of 24 solved) than was Problem A, which requires that the value of the result be changed (23 out of 24 solved) [$\chi^2(1, N = 24) = 18.38$, $p < .001$]. Problem C, which requires decomposition of a tight chunk, was solved significantly less often (18 out of 24 solved) than was Problem A, which requires decomposition of a loose chunk [$\chi^2(1, N = 24) = 4.18$, $p < .05$]. Moreover, Problem B was solved significantly less often than Problem C [$\chi^2(1, N = 24) = 4.78$, $p < .05$].

Solution times for Problems A, B, and C were analyzed next. For problems that were not solved, solution times were replaced with the upper time limit of 5 min. Because the distribution of solution times was highly skewed, we report median and quartiles and use ordinal tests to assess statistical significance. Table 1 shows median and lower and upper quartiles for Problems A, B, and C.

The medians, as well as the lower and upper quartiles, are consistent with the expected rank ordering of task difficulty (Problem B > Problem A, Problem C > Problem A). Single Wilcoxon tests were computed to assess the statistical significance of the differences. Problem A was solved significantly faster than Problem B ($n = 24$, $T = 1$, $p < .001$), and Problem A was solved significantly faster than Problem C ($n = 24$, $T = 22$, $p < .001$). In addition, Problem C was solved significantly faster than Problem B ($n = 24$, $T = 36$, $p < .05$).
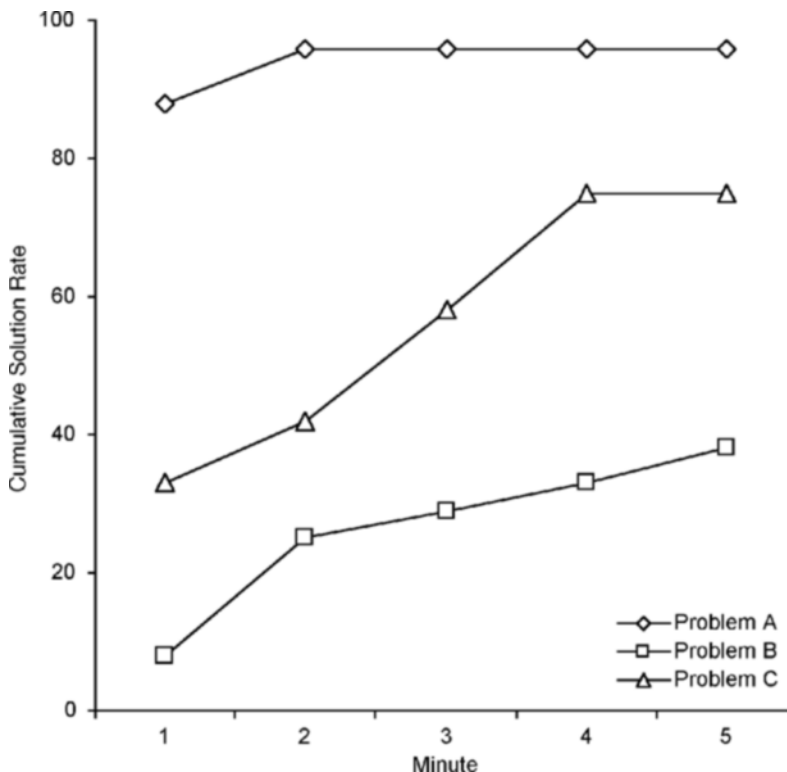

Figure 2. Cumulative solution rates for Problems A, B, and C.

Table 1
Solution Times (in Seconds) For Problems A, B, and C

Problem	Mdn	Lower Q	Upper Q
A	22	14	31
B	300	112	300
C	136	42	300

The observed differences in solution frequency and solution times are indistinguishable from those obtained in earlier experiments (Knoblich et al., 1999), during which no eye movements were recorded and subjects were not interrupted by recalibration procedures. In short, the eye movement recording procedure was not highly reactive in this situation.

Eye Movement Data

Only fixations that lasted longer than 100 msec were included in the data analysis. Before computing the proportion fixation time spent on each problem element, a median split was used to categorize the fixations into short and long. To trace changes in eye movements over time, the problem solving process was divided into three intervals of equal duration for each problem solver and each task. Because the total solution time varied, so did the absolute duration of the intervals. The relevant dependent measures were aggregated across participants for each interval.

Fixation duration. Figure 3 shows the results of the analysis of mean fixation duration across the three intervals.

Overall fixation duration was higher in Problems B and C than in Problem A. Mean fixation duration increased monotonously across intervals in Problems B and C, but remained constant after the second interval in Problem A. We computed a 3×3 analysis of variance (ANOVA) with problem (A, B, C) and interval (first, second, third) as within-subjects factors to assess the statistical significance of these differences. There was a significant main effect for problem [$F(2,44) = 3.49$, $p < .05$]. Post hoc comparisons showed that fixation duration for Problems B and C were significantly greater than for Problem A (both ps < .05). These results are consistent with the hypothesis that problem solvers encountered more impasses in Problems B and C. Furthermore, there was a significant main effect of interval [$F(2,44) = 8.58$, $p < .001$]. Mean fixation duration was longer in later intervals. There was no significant interaction between the two factors [$F(4,88) = 1.54$, $p = .20$]. Further analyses showed that there were also no significant differences in mean fixation duration between successful and unsuccessful problem solvers.

Fixation time on single elements during the initial phase. To determine whether there was a preference for certain problem elements in the initial phase of problem solving, we computed the proportion of fixation time spent on different problem elements during the first interval of problem solving. This measure was computed separately for fixations that were below and above the median of the mean fixation duration (i.e., for short and long fixations). If fixations were equally distributed across different elements of the display, one would expect that
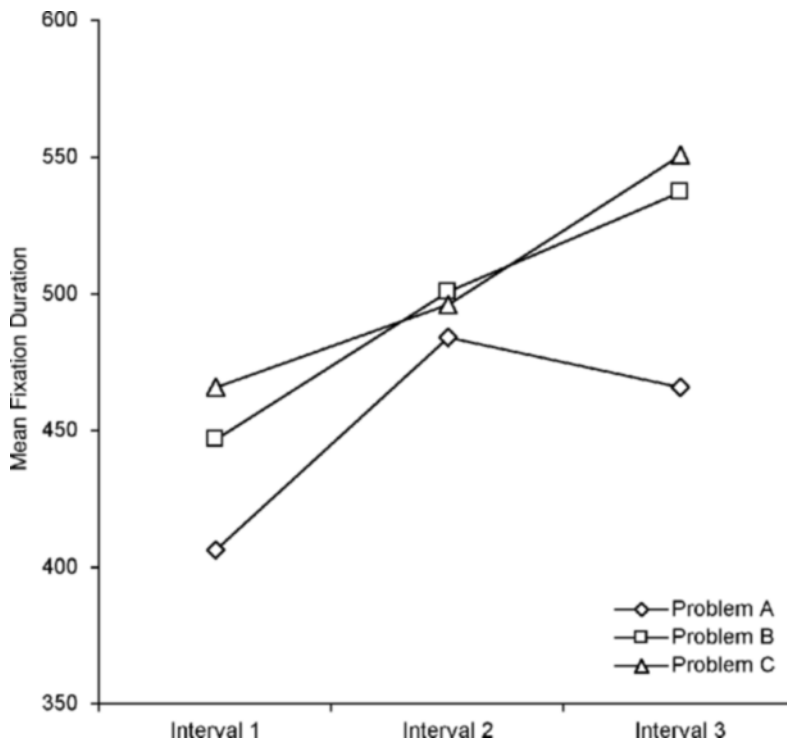

Figure 3. Mean fixation duration across intervals in Problems A, B, and C.

20% of the overall fixation time was spent on each element. Figure 4 shows the results (for the two operands, the arithmetic mean of the two percentages is reported).

For short fixations (see Panel A in Figure 4), the percentage of fixation time spent on the result was lower than the percentage time spent on all other elements. Consequently, the percentage fixation time was lower than the expected 20% for the result and slightly higher than that for the other problem elements. Accordingly, a 4×3 ANOVA with element (result, operands, plus sign, equal sign) and problem (A, B, C) as within-subjects factors resulted in a significant main effect for element [$F(3,63) = 2.82, p < .05$]. There was no significant main effect for problem ($p = .42$) and no significant interaction ($p = .26$).

For long fixations (see Panel B in Figure 4), the percentage of fixation time spent on the result was higher than the percentage time spent on the operands, which in turn was higher than the percentage time spent on the plus and the equal signs. The percentage of time spent on the plus and equal signs was clearly lower than the 20% that

would be expected if fixations were equally distributed across elements. Again, a 4×3 ANOVA with element (result, operands, plus sign, equal sign) and problem (A, B, C) as within-subjects factors revealed a significant main effect for element [$F(3,63) = 24.02, p < .001$] but no significant main effect for problem ($p = .42$). There was a marginally significant interaction [$F(6,126) = 2.06, p = .06$] that was due to the fact that the percentage of time spent on the plus sign was especially low in Problem A. Post hoc tests were conducted to confirm that the differences between the different conditions of the element factor were significant. This was true for the difference between the result and the operands conditions ($p < .01$) and the difference between the operands and the plus and equal sign conditions (both $ps < .001$). In short, there was a bias towards values for long fixations and no such bias for short fixations.

Fixation time on crucial elements across intervals. The next question is whether successful and unsuccessful problem solvers differed in the percentage of time spent
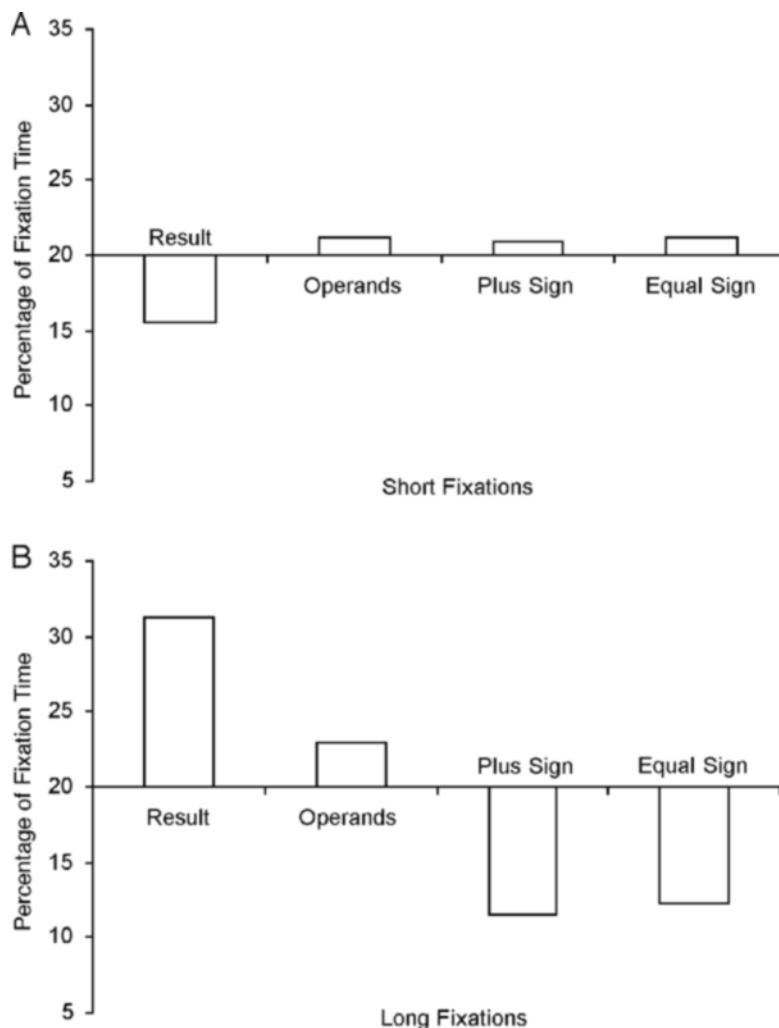

Figure 4. Percentage of fixation time spent on different problem elements during the initial phase of problem solving for short (A) and long (B) fixations.
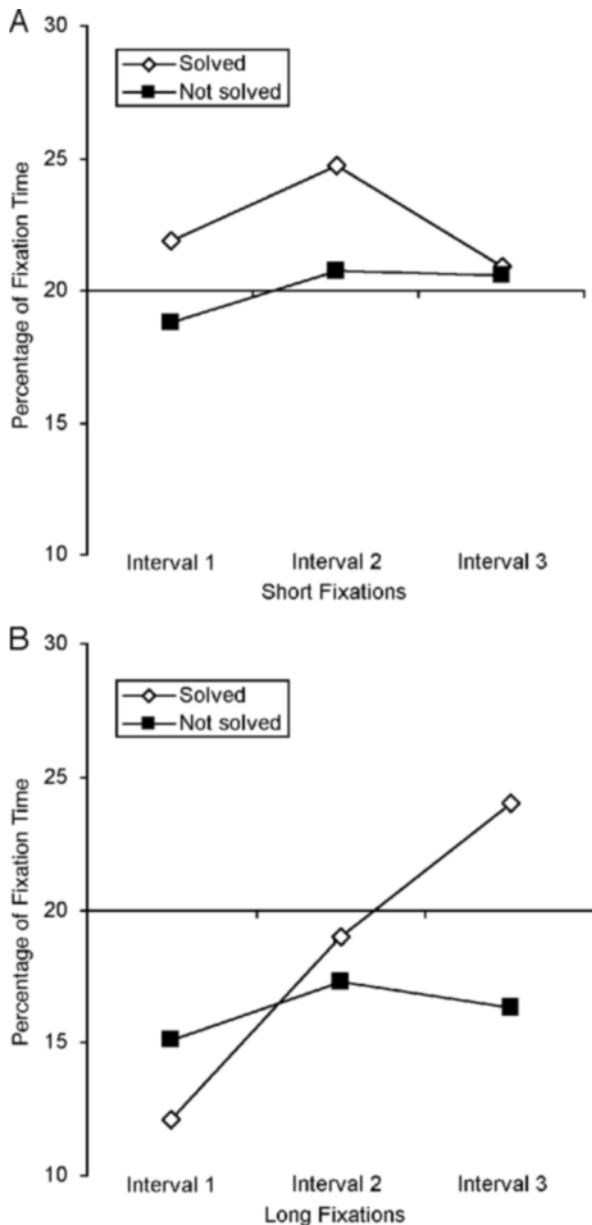
Figure 5. Percentage of fixation time spent on the operators in Problem B across intervals for successful and unsuccessful problem solvers and for short (A) and long (B) fixations.

on the crucial element across successive intervals. We first present the results from Problem B, in which the plus sign and the equal sign were the crucial problem elements.

For short fixations (left panel of Figure 5), the percentage of time spent on the operators was higher for successful problem solvers during the first two intervals. For unsuccessful problem solvers, the rate never deviated from the chance level of 20%. A 2×3 ANOVA with success (successful, unsuccessful) and interval (first, second, third) as within-subjects factors revealed no significant main effects or interactions. However, a contrast analysis confirms that the difference between successful and unsuc-

cessful problem solvers during the first two intervals was significant [$F(1,22) = 7.09, p < .05$].

For long fixations (right panel of Figure 5), the percentage of fixation time spent on the operators increased across intervals for successful problem solvers but not for unsuccessful problem solvers. Moreover, the percentage remained below chance level for unsuccessful problem solvers, but was above chance level in the third interval for successful problem solvers. A 2×3 ANOVA with success (successful, unsuccessful) and interval (first, second, third) as within-subjects factors revealed a significant main effect of interval [$F(2,44) = 12.35, p < .001$] and a significant interaction between success and interval [$F(1,22) = 7.91, p < .01$]. There was no significant main effect of success ($p = .28$).

In Problem C, the result was the crucial problem element. Figure 6 shows the percentage of fixation time spent on the result by successful and unsuccessful problem solvers across intervals. For short fixations, this percentage was consistently higher for unsuccessful problem solvers than for successful problem solvers. Moreover, it was always close to chance level for unsuccessful problem solvers, but consistently below chance level for successful problem solvers. However, a 2×3 ANOVA with success (successful, unsuccessful) and interval (first, second, third) as within-subjects factors revealed only a marginally significant main effect of success [$F(1,21) = 3.09, p = .09$]. There was no significant main effect of interval ($p = .78$) and no significant interaction of success and interval ($p = .87$).

For long fixations, the percentage of fixation time spent on the result was consistently above chance level in successful and unsuccessful problem solvers. There was no difference between the two groups during the first two intervals. However, during the third interval, the percentage fixation time increased dramatically for successful problem solvers and not at all for unsuccessful problem solvers. A 2×3 ANOVA with success (successful, unsuccessful) and interval (first, second, third) as within-subjects factors revealed a significant main effect for interval [$F(2,42) = 5.31, p < .01$] and a significant interaction between success and interval [$F(2,42) = 5.08, p < .05$] but no significant main effect for success ($p = .33$).

DISCUSSION

The experiment tested three predictions about eye movements derived from our theory of insight. The first prediction follows from the principle that insight problems produce impasses. During an impasse, people tend to sit and stare at the problem, so there should be fewer eye movements (i.e., longer fixation times). This effect should gradually increase throughout the problem solving process as more and more of the problems solvers in the sample enter impasses. We found this pattern for the two difficult problems, B and C, but not for the simpler problem, A (see Figure 3).

The second prediction follows from the principle that impasses are caused by inappropriate initial representations. In matchstick arithmetic, prior knowledge of arith-
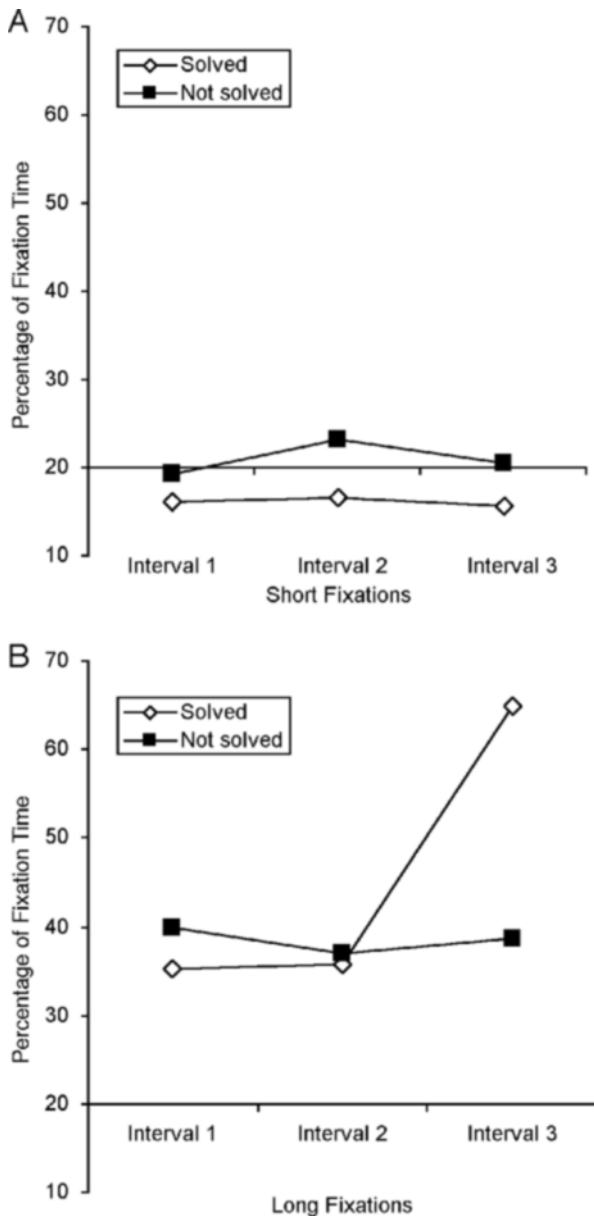
Figure 6. Percentage of fixation time spent on the result in Problem C across intervals for successful and unsuccessful problem solvers and for short (A) and long (B) fixations.

metic ought to bias people toward seeing values but not operators as variable problem elements. Hence, attention should be differentially allocated to the values during the initial phase of problem solving. This is the pattern we found (see Figure 4, Panel B).

The third prediction follows from the principle that impasses are resolved by relaxing inappropriate constraints and decomposing unhelpful perceptual chunks. Such revisions of the problem representation ought to result in shifts in attention allocation. In Problem B, we expected problem solvers to succeed because they finally relaxed the constraint

against manipulating the operators, which implies an increased attention to the operators toward the end of problem solving. This is what we found for successful problem solvers, but not for unsuccessful ones (see Figure 5, Panel B). In Problem C, we expected problem solvers to succeed by decomposing the chunk X into its component matchsticks, which should have led to increased attention to that symbol toward the end of problem solving, but only for successful problem solvers. This is what we found (see Figure 6, Panel B).

The conjunction of these three regularities is not predicted by other theories of insight. There are several principles that explain impasses, including functional fixedness, mental ruts, Einstellung, and insufficient progress (MacGregor et al., 2001). These principles all share the prediction that fixation time should increase throughout problem solving. However, because they do not explain how impasses are resolved, they cannot predict in detail how attention allocation will change at the moment of insight.

The principles of Gestalt restructuring, the variation–selection hypothesis, and gradual transformation of past experience all predict, at a general level, that attention should shift throughout a solution attempt. However, they do not predict any particular regularities in attention allocation. For example, the variation–selection hypothesis cannot explain why our experimental participants initially focused on the values in Problem B but later shifted their focus of attention to the operators, nor why this shift occurred only in successful problem solvers. Similarly, the gradual transformation principle cannot explain why attention shifted to the symbol X, specifically, toward the end of the successful solutions for Problem C.

Taken together, the three regularities we observed support our analysis of what happens when people solve arithmetic matchstick problems. The arithmetic format automatically activates prior knowledge about arithmetic, which brings with it mental habits and dispositions that are not useful for solving such problems, primarily the habit of regarding the values of the numbers as the variable elements of the problem and the operators and the equal sign as constant elements, and the disposition to see complex Roman numerals as indivisible entities. These habits and dispositions lead to impasses. When they are abandoned or overcome, attention allocation and hence processing shifts, the impasse is broken and the solution is quickly found. This analysis successfully accounts for both traditional performance measures (Knoblich et al., 1999) and the patterns in the eye movement recordings.

The success of the task specific analysis lends credence to the general representational change theory of insight from which it was derived. People encounter impasses on simple problems that they are competent to solve because the presentation of the problem triggers the retrieval or activation of prior knowledge that blocks the solution. Changes in the inappropriate initial representation allow previously unheeded possibilities to come to mind; this is the moment of insight. However, there is still room for dis-

agreement about the exact nature of the processes of representational change. Kaplan and Simon (1990) emphasize the incompleteness of the initial representation and the importance of invariant properties. This hypothesis has a high degree of face validity with respect to their experimental task, the *mutilated checkerboard problem.* However, it is not clear which invariant properties there are in our matchstick problems, nor in the classical insight problems such as the two-string problem or the nine-dot problem. We prefer the hypothesis that initial representations are inappropriate or misleading rather than incomplete, and thus have to be deactivated or inhibited rather than extended or elaborated, but the available data leave the issue open.

The present study demonstrates the power of eye movement recordings. Traditional performance measures like solution time and solution rate could not have revealed the structure of the participants' attention allocation in nearly the same amount of detail. We expect that extensive use of eye movement recordings will enable insight researchers to empirically differentiate between alternative mechanisms of representational change.

REFERENCES

ANDERSON, J. R., & LEBIERE, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

BALLARD, D. H., HAYHOE, M. M., POOK, P. K., & RAO, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences*, **20**, 723-767.

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

DUNCKER, K. (1945). On problem-solving. *Psychological Monographs*, **68** (Whole No. 270).

JUST, M. A., & CARPENTER, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, **8**, 441-480.

KAPLAN, C. A., & SIMON, H. A. (1990). In search of insight. *Cognitive Psychology*, **22**, 374-419.

KATONA, G. (1940). *Organizing and memorizing: Studies in the psychology of learning and teaching.* New York: Columbia University.

KEANE, M. (1989). Modelling problem solving in Gestalt "insight" problems. *Irish Journal of Psychology*, **10**, 201-215.

KNOBLICH, G., OHLSSON, S., HAIDER, H., & RHENIUS, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1534-1556.

KNOBLICH, G., & RHENIUS, D. (1995). *Zur Reaktivität Lauten Denkens beim komplexen Problemlösen* [The reactivity of thinking aloud during complex problem-solving]. *Zeitschrift für Experimentelle Psychologie*, **42**, 419-454.

KNOBLICH, G., & WARTENBERG, F. (1998). *Unbemerkte Lösungshinweise begünstigen Veränderungen der Problemrepräsentation* [Unnoticed hints facilitate representational change in problem solving]. *Zeitschrift für Psychologie*, **206**, 207-234.

KÖHLER, W. (1924). *Die physischen Gestalten in Ruhe und im stationären Zustand* [The physical shapes in rest and stationary state]. Erlangen, Germany: Verlag der philosophischen Akademie.

KÖHLER, W. (1925). *The mentality of apes.* New York: Livewright.

LUCHINS, A. S., & LUCHINS, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung.* Eugene: University of Oregon Press.

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

MAIER, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, **12**, 181-194.

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

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

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

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

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

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

RAYNER, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, **85**, 618-660.

RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.

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

SCHOOLER, J. W., & MELCHER, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97-133). Cambridge, MA: MIT Press.

SCHOOLER, J. W., OHLSSON, S., & BROOKS, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, **122**, 166-183.

SIMON, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and cosmos* (pp. 22-40). Pittsburgh: University of Pittsburgh Press.

SIMONTON, D. K. (1988). *Scientific genius: A psychology of science.* New York: Cambridge University Press.

SIMONTON, D. K. (1995). Foresight in insight? A Darwinian answer. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 465-494). Cambridge, MA: MIT Press.

SMITH, S. M. (1995). Getting into and out of mental ruts: A theory of fixation, incubation, and insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 229-251). Cambridge, MA: MIT Press.

SMITH, S. M., WARD, T. B., & FINKE, R. A. (EDS.) (1995). *The creative cognition approach.* Cambridge, MA: MIT Press.

STERNBERG, R. J., & DAVIDSON, J. E. (EDS.) (1995). *The nature of insight.* Cambridge, MA: MIT Press.

WEISBERG, R. W. (1986). *Creativity: Genius and other myths.* New York: W. H. Freeman.

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

WEISBERG, R. W., & SULS, J. M. (1973). An information-processing model of Duncker's Candle Problem. *Cognitive Psychology*, **4**, 255-276.

WERTHEIMER, M. (1959). *Productive thinking.* New York: Harper.