Memory & Cognition
1987, 15(3), 238-246

Intuition in insight and noninsight problem solving

JANET METCALFE
Indiana University, Bloomington, Indiana

and

DAVID WIEBE
University of British Columbia, Vancouver, British Columbia, Canada

People's metacognitions, both before and during problem solving, may be of importance in motivating and guiding problem-solving behavior. These metacognitions could also be diagnostic for distinguishing among different classes of problems, each perhaps controlled by different cognitive processes. In the present experiments, intuitions on classic insight problems were compared with those on noninsight and algebra problems. The findings were as follows: (1) subjective feeling of knowing predicted performance on algebra problems but not on insight problems; (2) subjects' expectations of performance greatly exceeded their actual performance, especially on insight problems; (3) normative predictions provided a better estimate of individual performance than did subjects' own predictions, especially on the insight problems; and, most importantly, (4) the patterns-of-warmth ratings, which reflect subjects' feelings of approaching solution, differed for insight and noninsight problems. Algebra problems and noninsight problems showed a more incremental pattern over the course of solving than did insight problems. In general, then, the data indicated that noninsight problems were open to accurate predictions of performance, whereas insight problems were opaque to such predictions. Also, the phenomenology of insight-problem solution was characterized by a sudden, unforeseen flash of illumination. We propose that the difference in phenomenology accompanying insight and noninsight problem solving, as empirically demonstrated here, be used to define insight.

The rewarding quality of the experience of insight may be one reason why scientists and artists alike are willing to spend long periods of time thinking about unsolved problems. Indeed, creative individuals often actively seek out weaknesses in theoretical structures, areas of unresolved conflict, and flaws in conceptual systems. This tolerance and even questing for problems carries the risk that a particular problem may have no solution or that the investigator may be unable to uncover it. For instance, it was thought for many years that Euclid's fifth postulate might be derivable even though no one was able to derive it (see Hofstadter, 1980). The payoff for success, however, is the often noted "discovery" experience for the individual and (perhaps) new knowledge structures for the culture. This special mode of discovery may be qualitatively different from more routine analytical thinking.

Bergson (1902) differentiated between an intuitive mode of inquiry and an analytical mode. Many other theorists have similarly emphasized the importance of a method of direct apperception, variously called restructuring, intuition, illumination, or insight (Adams, 1979; Bruner, 1966; Davidson & Sternberg, 1984; Dominowski, 1981; Duncker, 1945; Ellen, 1982; Gardner, 1978; Koestler, 1977; Levine, 1986; Maier, 1931; Mayer, 1983; Polya, 1957; Sternberg, 1986; Sternberg & Davidson, 1982; Wallas, 1926). Polanyi (1958) noted:

> We may describe the obstacle to be overcome in solving a problem as a "logical gap," and speak of the width of the logical gap as the measure of the ingenuity required for solving the problem. "Illumination" is then the leap by which the logical gap is crossed. It is the plunge by which we gain a foothold at another shore of reality. (p. 123)

Sternberg (1985) said that "significant and exceptional intellectual accomplishment—for example, major scientific discoveries, new and important inventions, and new and significant understandings of major literary, philosophical, and similar work—almost always involve [sic] major intellectual insights" (p. 282). Arieti (1976) stated that "the experience of aesthetic insight—that is, of creating an aesthetic unity—is a strong emotional experience.... The artist feels almost as if he had touched the universal" (p. 186). Although these major insights are of crucial importance both to the person and to the culture, their unpredictable and subjective nature presents difficulties for rigorous investigation. Sternberg and

This research was supported by Natural Sciences and Engineering Research Council of Canada Grant A-0505 to the first author. Thanks go to Judith Goldberg and Leslie Kiss for experimental assistance. Thanks also go to W. J. Jacobs. Please send correspondence about this article to Janet Metcalfe, Department of Psychology, Indiana University, Bloomington, IN 47405.

Davidson (1982) suggested that solving small insight puzzles may serve as a model for scientific insight. We shall adopt this approach in the present paper.

Despite the importance attributed to the process of insight, there is little empirical evidence for it. In fact, Weisberg and Alba (1981a) claimed correctly that there was no evidence whatsoever (see also Weisberg & Alba, 1981b, 1982). Since that time, two studies investigating the metacognitions that precede and accompany insight problem solving have provided some data favoring the construct. In the first study (Metcalfe, 1986a), feeling-of-knowing performance was compared on classical insight problems and on general information memory questions (Nelson & Narens, 1980). In the problem-solving phase of the study, subjects were given insight problems to rank order in terms of the likelihood of solution. On the memory half of the study, trivia questions that subjects could not answer immediately (e.g., "What is the name of the villainous people who lived underground in H. G. Wells's book *The Time Machine*?") were ordered in terms of the likelihood of remembering the answers on the second test. The memory part of the study was much like previous feeling-of-knowing experiments on memory (e.g., Gruneberg & Monks, 1974; Hart, 1967; Lovelace, 1984; Nelson, Leonesio, Landwehr, & Narens, 1986; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983). Metcalfe found that the correlation between predicted solution and actual solution was not different from zero for the insight problems, although this correlation, as in other research, was substantial for the memory questions. Metcalfe interpreted these data as indicating that insightful solutions could not be predicted in advance, which would be expected if insight problems were solved by a sudden "flash of illumination." However, the data may have resulted from a difference between problem solving in general and memory retrieval, rather than a difference between insight and noninsight problem-solving processes.

In a second study (Metcalfe, 1986b), subjects were instructed to provide estimates of how close they were to the solutions to problems every 10 sec during the problem-solving interval. These estimates are called feeling-of-warmth (Simon, Newell, & Shaw, 1979) ratings. If the problems were solved by what subjectively is a sudden flash of insight, one would expect that the warmth ratings would be fairly low and constant until solution, at which point they would jump to a high value. This is what was found in the experiment. On 78% of the problems and anagrams for which subjects provided the correct solution, the progress estimates increased by no more than 1 point, on a 10-point scale, over the entire solution interval. On those problems for which the wrong answer was given, however, the warmth protocols showed a more incremental pattern. Thus, it did not appear to be the case that there were no circumstances at all under which an incremental pattern would appear. It appeared with incorrect solutions. However, in that study, the incremental pattern may have been attributable to a special decision-making strategy, rather than to an incremental problem-solving process. Thus, whether noninsight problems show a warmth pattern different from that of insight problems is still unclear.

A straightforward comparison of the warmth ratings produced during solution of insight and noninsight problems is, therefore, important and has not been attempted previously. Simon (1977, 1979) provided several models that apply to incremental problems such as algebra, chess, and logic problems. Basically, Simon et al. (1979) proposed that people are able to use a directed-search strategy in problem solving (as opposed to an exhaustive search through all possibilities, which in many cases would be impossible) because they are able to compare their present state with the goal state. If a move makes the present state more like a goal state (i.e., if the person gets "warmer"), that move is taken. Simon et al. (1979) provided several think aloud protocols that suggest that this "functional" or "means-end" analysis of reducing differences can be applied to a wide range of analytical problems. They noted that the Logic Theorist (a computer program that uses this heuristic) "can almost certainly transfer without modification to problem solving in trigonometry, algebra, and probably other subjects such as geometry and chess" (p. 157). If this monitoring process is guiding human problem solving, then the warmth ratings should increase to reflect subjects' increasing nearness to the solutions. Of course, if insight problems are solved by some nonanalytical, sudden process, as previous research suggests, we would expect to find a difference, depending on problem type, in the warmth protocols.

The experiments described below explored the metacognitions exhibited by subjects on insight and on noninsight problems. Experiment 1 compared warmth ratings during the solution of insight problems with those produced during the solution of noninsight problems. The noninsight problems were the type that have been analyzed and modeled by programs that use a functional analysis. Thus, we expected to find that subjects' warmth ratings would increment gradually over the course of the problem-solving interval. We expected that warmth ratings on the insight problems would, in contrast, increase rapidly only when the solution was given. Experiment 2 used algebra problems rather than the multistep problems that have been modeled with search-style programs. Algebra problems may be more characteristic of the sorts of problems people solve daily than are (at least some of) the multistep problems used in Experiment 1. As noted above, however, because the means-end search strategy should be applicable to algebra problems, incremental warmth protocols were expected. For these reasons, as well as their availability, algebra problems were worth investigating. In addition to examining warmth ratings during the course of problem solving, Experiment 2 also investigated other predictors of performance: subjects' feeling-of-knowing rankings, normative predictors of performance, and subjective estimations of the likelihood of success. We expected that noninsight problems would show more incremental warmth protocols than would insight problems. We also expected that people would have more accurate metacognitions (about how well they would be able to solve problems and which problems they would be able to solve) for the noninsight than for the insight problems.

EXPERIMENT 1

Method

Subjects. Twenty-six volunteers were paid $4 for a 1-h session of problem solving. Seven of these subjects either produced no correct answers on one of the insight or the noninsight problems or produced correct answers immediately, so that no warmth protocols could be obtained for the solution interval. Thus, 19 subjects produced usable data.

Materials. Ten problems, provided on 3×5 in. cards, were given in random order to the subjects for solution one at a time. Half of these problems were noninsight problems, and half were insight problems. The noninsight problems were designated as such because past literature had labeled them as multistep problems or because they had been analyzed by incremental or search models such as those of Karat (1982) or Simon (1977, 1979). The noninsight problems are reproduced in Appendix A. The insight problems were chosen because they had been considered to be insight problems by other authors or by the sources from which they were taken. However, we felt free to eliminate problems that in our previous experiments (Metcalfe, 1986a, 1986b) had been designated by subjects as "grind-out-the-solution" problems rather than insight problems. Our criterion for calling a problem an insight problem was not well defined. This lack of definition may well be one reason that research on insight has progressed so slowly. We shall return to this point in our conclusion. The insight problems we used are reproduced in Appendix B.

Procedure. The subjects were told that they would be asked to solve a number of problems, one at a time. Once they had the answer, they were to write it down so that the experimenter could ascertain whether it was right or wrong. If the experimenter had any doubt about the correctness of the answer, she asked the subject for clarification before proceeding to the next problem. During the course of solving, the subjects were asked to provide warmth ratings to indicate their perceived nearness to the solution. These ratings were marked by the subject with a slash on a 3-cm visual analogue scale on which the far left end was "cold," the far right end was "hot," and intermediate degrees of warmth were to be indicated by slashes in the middle range. Altogether, there were 40 lines that could be slashed for each problem (to allow for the maximum amount of time that a subject was permitted to work on a given problem); these lines were arranged vertically on an answer sheet. The subjects were told to put their first rating at the far left end of the visual analogue scale. They then worked their way down the sheet marking warmth ratings at 15-sec intervals, which were indicated by a click given by the experimenter. Because it requires less attention on the part of the subject, is nonsymbolic, and apparently is less distracting and intrusive, this visual-analogue-scale technique for assessing warmth is superior to the Metcalfe (1986b) technique of writing down numerals.

Results

The probability level of $p \leq .05$ was chosen for significance. The increments in the warmth ratings were assessed in two ways. First, the angle subtended from the first rating to the last rating before the rating given with the answer in a particular protocol for a particular problem was measured. (We refer to this henceforth as the "angular warmth"). Second, the difference between the first slash or rating and the last rating before the rating given with the answer was measured. (We refer to this as the "differential warmth"). These two methods can yield different results because angular warmth, unlike differential warmth, varies according to the total time spent solving the problem. For example, consider two protocols, both of which start at the far left end of the scale and both of which have the warmth immediately before the answer given as a slash in the exact center of the scale. Let the first protocol have a total time of 1 min and the second a total time of 2 min. When these two are ranked according to differential warmth, they will be tied, or be considered to be equally incremental. When they are ranked according to angular warmth, however, the first protocol will be said to be more incremental than the second. Thus, the differential warmth measure considers the total solving time (whatever it is) to be the unit of analysis, whereas the angular warmth measure gives an indication of the increment in warmth per unit of real time. We could not decide which method was more appropriate, so we used both.

The correctly solved problems were separately rank ordered from greatest to least on each of the angular warmth and the differential warmth measures, for each subject. Then a Goodman and Kruskal gamma correlation (see Nelson, 1984, 1986), comparing the rank orderings of the increment in warmth (going from most incremental to least) and problem type was computed. These gammas were treated as summary data scores for each subject. A positive correlation (which is what was expected) indicates that the noninsight problems tended to have more incremental warmth protocols than did the insight problems. The overall correlation on the angular warmth measure was .26, which is significantly different from zero [$t(18) = 2.02$, $MSe = .32$]. The overall correlation on the differential warmth measure was .23, which was also significantly different from zero [$t(18) = 1.63$, $MSe = .37$, by a one-tailed test].

Thus, the warmth protocols of the insight problems in Experiment 1 showed a more sudden achievement of solution than did those of the noninsight problems. This is precisely what was expected given that the insight problems involved sudden illumination and the noninsight problems did not.

EXPERIMENT 2

Method

Subjects. Seventy-three University of British Columbia students in introductory psychology participated in exchange for a small bonus course credit. To allow assessment of performance on the feeling-of-knowing tasks detailed below, it was necessary that the subjects correctly solve at least one insight and one algebra problem, and that they miss at least one insight and one algebra problem. Twenty-one subjects failed to get at least one algebra or one insight problem correct, and so were dropped from the analyses. Four subjects got all the algebra problems correct and were dropped. This left 48 subjects who provided usable feeling-of-knowing data.

For warmth-rating data to be usable, it was necessary that the subjects get at least one insight and one algebra problem correct with at least three warmth ratings. Thirty-nine subjects provided usable data for this analysis.

Materials. The materials were classical insight problems (reproduced in Appendix B) and algebra problems selected from a high school algebra textbook (reproduced in Appendix C). The insight problems were selected, insofar as possible, to require little cognitive work other than the critical insight. Weisberg and Alba (1981b) argued against the idea of insight because they found that providing the clue or "insight" considered necessary to solve a problem did not ensure problem solution. Sternberg and Davidson (1982) pointed out that this failure of the clue to result in immediate problem solution may have occurred not because there was no process of insight, but rather because there were a number of additional processes involved in solving the problem as well as insight. In an attempt to circumvent such additional processes, we tried to use problems that were minimal.

Procedure and Design. The subjects were shown a series of insight or algebra problems, one at a time, randomly ordered within insight or algebra problem-set block. If they knew the answer to the problem, either from previous experience or by figuring it out immediately, the problem was eliminated from the test set. Once five unsolved problems (either insight or algebra, depending on order of presentation condition) had been accumulated, the experimenter arranged in a circle the 3×5 in. index cards on which the problems were typed and asked the subjects to rearrange them into a line going from the problem they thought they were most likely to be able to solve in a 4-min interval to that which they were least likely to be able to solve. This ranking represents the subjects' feelings that they will know (or feeling-of-knowing) ordering. The five cards were reshuffled, and the subjects were asked to assess the probability that they could solve each problem. The cards were shuffled again and then presented one at a time for solution. Every 15 sec during the course of solving, the subjects were told to indicate their feeling of warmth (i.e., their perceived closeness to solution) by putting a slash through a line that was 3 cm long, as in Experiment 1. The subjects were not told explicitly to anchor the first slash at the far left of the scale, but they tended to do so. Altogether, there were 17 lines that could be slashed for each problem. The subjects continued through the set of five test problems until they had either written a solution or exhausted the time on each. Then the procedure was repeated with the other set of problems (either insight or algebra). The order of problem set (insight or algebra) was counterbalanced across subjects. The subjects were tested individually in 1-h sessions.

Results

Warmth ratings. The gammas computed on the angular warmth measures indicated that the insight problems showed a less incremental slope than did the algebra problems: Mean $G = .35$, which is significantly greater than zero [$t(37) = 3.10$, $MSe = .49$]. Gammas computed on the differential warmth measure showed the same pattern: Mean $G = .32$, which is also significantly greater than zero [$t(37) = 2.56$, $MSe = .58$].

Figure 1 provides a graphical representation of the subjects' warmth values for insight and algebra problems during the minute before the correct solution was given. The histograms in Figure 1 contain data from all subjects who had ratings in the specified intervals. To convert the visual analogue scale to a numerical scale, the 3-cm rating lines were divided into seven equal regions, and a slash occurring anywhere within one of these regions was given the appropriate numerical warmth value. Thus, ratings of 7 could occur before a solution was given because the subjects could, and did, provide ratings that were almost, but not quite, at the far right end of the scale. The trends of the distributions, over the last minute of solving time, going from the bottom to the top panel in Figure 1, tell the same story as the angular and differential warmth measures: There was a gradual increment in warmth with algebra problems but little increasing warmth with the insight problems.

Feeling of knowing on ranks. A Goodman and Kruskal gamma correlation was computed between the rank ordering given by the subject and the response (correct or incorrect) on each problem, for each of the two sets of problems. Then an analysis of variance was performed on these scores; the factors were order of presentation of problem block (either algebra first or insight first—between subjects) and problem type (algebra or insight—within subjects). There was a significant difference in gamma between the algebra problems (mean $G = .40$) and the insight problems (mean $G = .08$) [$F(1,46) = 6.46$, $MSe = .77$]. The correlation on the algebra problems was greater than zero [$t(46) = 4.6$, $MSe = .36$], whereas the correlation for the insight problems was not [$t(46) = .8$, $MSe = .47$]. This latter result replicates Metcalfe (1986a). Thus, it appears that the subjects fairly accurately predicted which algebra problems they would be able to solve later, but were unable to predict which insight problems they would solve.

Feeling of knowing on probabilities. The problems in each set were ranked according to the stated probability that they would be solved, and another gamma was computed on the data so arranged. Because the correlation cannot be computed if the identical probabilities are given for all problems (in either set), 4 subjects had to be eliminated from this analysis, leaving 44 subjects. Because there could be ties in the probability estimates, and because there could be some inconsistency between the rankings and the estimates, the results are not identical to those presented above. As before, the difference between the algebra problems (mean $G = .40$) and the insight problems (mean $G = .15$) was significant [$F(1,42) = 2.8$, $MSe = .99$ one-tailed]. The correlation for the algebra problems was significantly greater than zero [$t(42) = 3.82$, $MSe = .48$], whereas the correlation for the insight problems was not [$t(42) = 1.2$, $MSe = .65$]. This analysis is consistent with the analysis conducted on the ranks.

Calibration. To compare the subjects' overall ability (Lichtenstein, Fischoff, & Phillips, 1982) to predict how well they would perform on the insight versus the algebra problems, the mean value was computed for the five probability estimates (one for each problem). This mean was compared with the actual proportion of problems that each subject solved correctly. Both the predicted performance

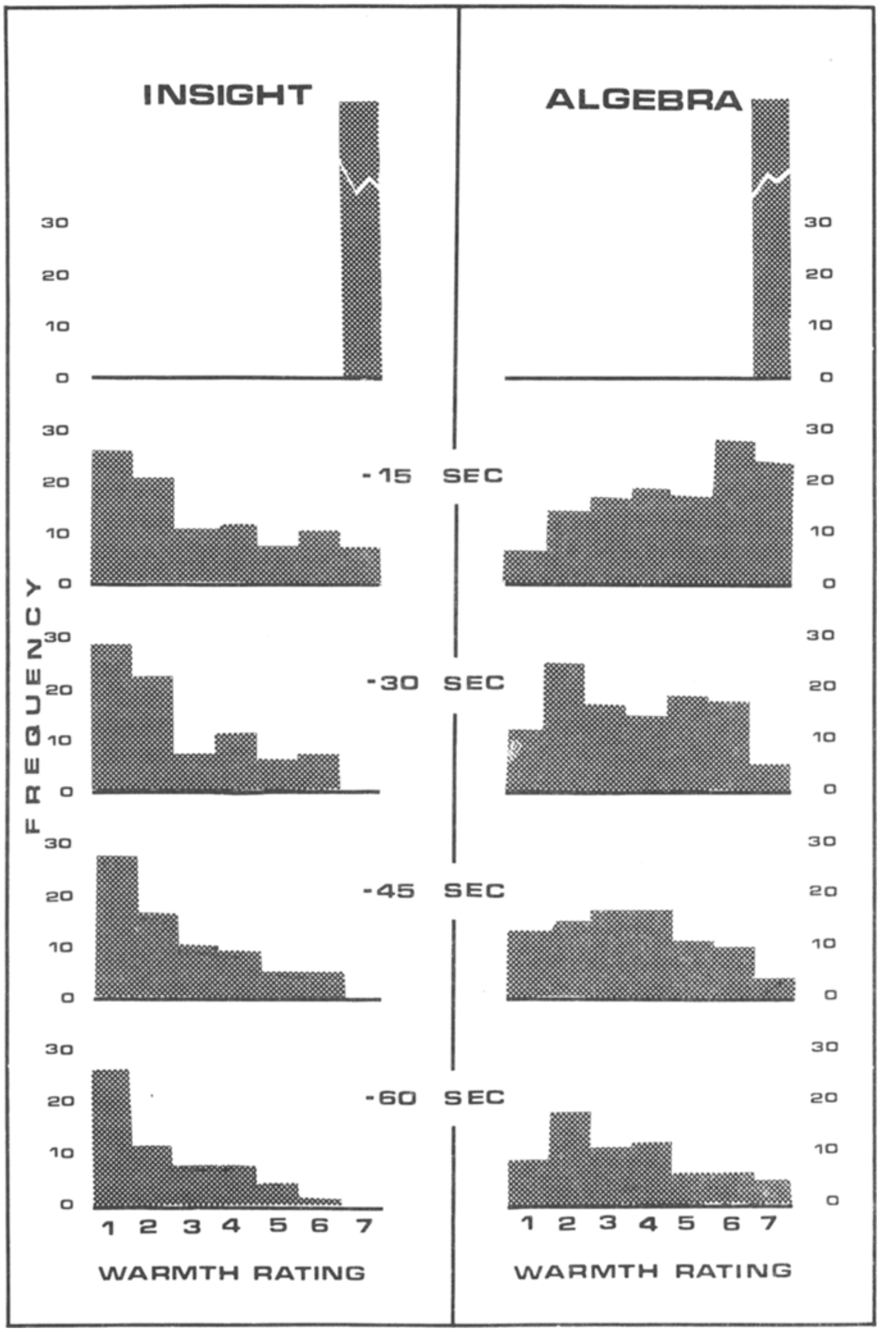


Figure 1. Frequency histograms of warmth ratings for correctly solved insight and algebra problems in Experiment 2. The panels, from bottom to top, give the ratings 60, 45, 30, and 15 sec before solution. As shown in the top panel, a 7 rating was always given at the time of solution.

and the actual performance were better on the algebra problems than on the insight problems [$F(1,46) = 54.67$, $MSe = .11$]. Previous research had shown that subjects overestimated their ability more on insight problems than on memory questions (Metcalfe, 1986a), and we thought they they would perhaps overestimate more on the insight problems than on the algebra problems. Predicted performance on the insight problems was .59, whereas actual performance was only .34. Predicted performance was .73 on the algebra problems, whereas actual performance was .55. The interaction showing that there was greater overestimation on the insight than on the algebra problems was significant [$F(1,46) = 3.18$, $MSe = .08$, one-tailed]. This result is fairly weak. Not only is the interaction significant only by a one-tailed test, but also, insofar as the actual performance differed between the insight and algebra problems, the interaction involving predicted performance could be eliminated by changing the scale. Despite these hedges, the result suggests that people may overestimate their ability more on insight than on algebra problems. None of the interactions with order of set was significant.

Personal versus normative predictions. We looked at normative predictions because it was possible that there was no, or a very diffuse, underlying difficulty structure (or a restricted range in the probability correct) with the insight problems, and hence the zero feeling-of-knowing correlations could simply reflect that lack of structure, or range. In addition, there is the interesting possibility that the normative predictions of problem difficulty are more accurate at predicting individual behavior in particular situations than are subjects' self-evaluations. Nelson et al. (1986) found such an effect with memory retrieval. If this were the case in problem solving as well, then the experimenter would in theory be able to predict better than a person him- or herself whether that person would solve a particular problem.

The problems were ranked ordered in terms of their difficulty by computing across subjects the probability of solution for each. Although ideally difficulty should have been computed from an independent pool of subjects, this was not cost effective. Thus, there is a small artificial correlation induced in this ranking because a subject's own results made a 2.1%, rather than a 0% contribution to the difficulty ranking. To see whether normative ranking was better than subjective judgment as a predictor of individual problem-solving performance, two gammas were compared. The first was based on the normative ranking against the individual's performance, and the second was based on the subject's own feeling-of-knowing rank ordering against his or her performance. The normative probabilities were a much better predictor of subjects' individual performance than their own feelings of knowing. The normative correlation for the insight problems was .77; for the algebra problems, it was .60. These correlations indicate that there was sufficient range in the difficulty of the problems (both insight and algebra) that overall frequency correct was a good predictor of individual performance. The zero feeling-of-knowing correlation, discussed earlier, is therefore probably not attributable to a restricted range of insight-problem difficulty. The interaction between own versus normative gammas as a function of problem type was significant [$F(1,46) = 10.13$, $MSe = 1.13$]. Table 1 gives the means. The idea that subjects may have privileged access to idiosyncratic information that makes them especially able to predict their own performance was overwhelmingly wrong in this experiment.

Table 1
Mean Gamma Correlations Between Personal and Normative Predictions and Actual Performance for Insight and Algebra Problems in Experiment 2

	Type of Prediction	
	Personal	Normative
Insight	.08	.77
Algebra	.40	.60

DISCUSSION

This study shows that there is an empirically demonstrable distinction between problems that people have thought were insight problems and those that are generally considered not to require insight, such as algebra or multistep problems. The above experiments showed that people's subjective metacognitions were predictive of performance on the noninsight problems, but not on the insight problems. In addition, the warmth ratings that people produced during noninsight problem solving showed a more incremental pattern, in both experiments, than did those problems that were preexperimentally designated as involving insight. These findings indicate in a straightforward manner that insight problems are, at least subjectively, solved by a sudden flash of illumination; noninsight problems are solved more incrementally.

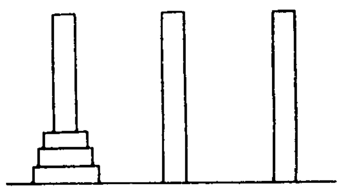
A persistent problem has blocked the study of the process of insight: How can we ascertain when we are dealing with an insight problem? Let us now propose a solution. Given that the warmth protocols differentiate between problems that seem to be insight problems and those that do not, we may use the warmth protocols themselves in a diagnostic manner. If we find problems (or indeed problems for particular individuals) that are accompanied by step-function warmth protocols during the solution interval, we may define those problems as being insight problems for those people. Thus, we propose that insight be defined in terms of the antecedent phenomenology that may be monitored by metacognitive assessments by the subject. Adopting this solution may have interesting (although as yet unexplored) consequences. Perhaps the underlying processes involved in solving an insight problem are qualitatively different from those involved in solving a noninsight problem. It may (or may not) be that contextual or structural novelty is essential for insight. Perhaps there is a class of problems that provoke insights for all people. But perhaps insight varies with the level of skill within a particular problem-solving domain. If so, we might be able to use the class of problems that provoke insight for an individual to denote the individual's conceptual development in the domain in question. Perhaps this person-problem interaction will provide some optimal difficulty level for motivating a person and therefore have pedagogical consequences. Insight problems may be especially challenging to people, and their solution distinctly pleasurable. Of course, many other possibilities present themselves for future consideration. The process of insight has heretofore been virtually opaque to scientific scrutiny. Differentiating insight problems from other problems by the phenomenology that precedes solution may facilitate illumination of the process of insight.

REFERENCES

Adams, J. L. (1979). *Conceptual blockbusting*. New York: Norton.
Arieti, S. (1976). *Creativity: The magic synthesis*. New York: Basic Books.
Bergson, H. (1902). *An introduction to metaphysics*. New York: Putnam.
Bruner, J. (1966). *On knowing: Essays for the left hand*. Cambridge: Harvard University Press.
Davidson, J. E., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly*, **28**, 58-64.
deBono, E. (1967). *The use of lateral thinking*. New York: Penguin.
deBono, E. (1969). *The mechanism of mind*. New York: Penguin.
Dominowski, R. L. (1981). Comment on an examination of the alleged role of 'fixation' in the solution of 'insight' problems. *Journal of Experimental Psychology: General*, **110**, 199-203.
Duncker, K. (1945). On problem solving. *Psychological Monographs*, **58**(5, Whole No. 270).
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.
Fixx, J. F. (1972). *More games for the superintelligent*. New York: Popular Library.
Gardner, M. (1978). *Aha! Insight*. New York: Freeman.
Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing in cued recall. *Acta Psychologica*, **38**, 257-265.
Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning & Verbal Behavior*, **6**, 685-691.
Hofstadter, D. R. (1980). *Gödel, Escher, Bach: An eternal golden braid*. New York: Vintage Books.
Karat, J. (1982). A model of problem solving with incomplete constraint knowledge. *Cognitive Psychology*, **14**, 538-559.
Koestler, A. (1977). *The act of creation*. London: Picadoo.
Levine, M. (1986). *Principles of effective problem solving*. Unpublished manuscript, State University of New York at Stonybrook.
Lichtenstein, S., Fischoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state to the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 756-766.
Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs*, **54**(6, Whole No. 248).
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.
Mayer, R. E. (1983). *Thinking, problem solving, cognition*. New York: Freeman.
Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 288-294.
Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 623-634.
Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling of knowing prediction. *Psychological Bulletin*, **95**, 109-133.
Nelson, T. O. (1986). ROC curves and measures of discrimination accuracy: A reply to Swets. *Psychological Bulletin*, **100**, 128-132.
Nelson, T. O., Leonesio, R. J., Landwehr, R. S., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing vs. the normative feeling of knowing vs. base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 279-287.
Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 279-288.
Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, **19**, 338-368.
Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago: University of Chicago Press.
Polya, G. (1957). *How to solve it*. Princeton, NJ: Princeton University Press.
Restle, F., & Davis, J. H. (1962). Success and speed of problem solving by individuals and groups. *Psychological Review*, **69**, 520-536.
Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 39-54.
Simon, H. (1977). *Models of discovery*. Dordrecht, The Netherlands: Reidel.
Simon, H. (1979). *Models of thought*. New Haven: Yale University Press.
Simon, H., Newell, A., & Shaw, J. C. (1979). The process of creative thinking. In H. Simon (Ed.), *Models of thought* (pp. 144-174). New Haven: Yale University Press.
Sternberg, R. J. (1985). *Beyond IQ*. Cambridge, MA: Cambridge University Press.
Sternberg, R. J. (1986). *Intelligence applied*. San Diego: Harcourt, Brace, Javonovich.
Sternberg, R. J., & Davidson, J. E. (1982, June). The mind of the puzzler. *Psychology Today*, **16**, 37-44.
Travers, K. J., Dalton, L. C., Bruner, V. F., & Taylor, A. R. (1976). *Using advanced algebra*. Toronto: Doubleday.
Wallas, G. (1926). *The art of thought*. New York: Harcourt.
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.
Weisberg, R. W., & Alba, J. W. (1982). Problem solving is not like perception: More on gestalt theory. *Journal of Experimental Psychology: General*, **111**, 326-330.

APPENDIX A
Incremental Problems

1. If the puzzle you solved before you solved this one was harder than the puzzle you solved after you solved the puzzle you solved before you solved this one, was the puzzle you solved before you solved this one harder than this one? (Restle & Davis, 1962)
2. Given containers of 163, 14, 25, and 11 ounces, and a source of unlimited water, obtain exactly 77 ounces of water. (Luchins, 1942)
3. Given state:



Goal state:

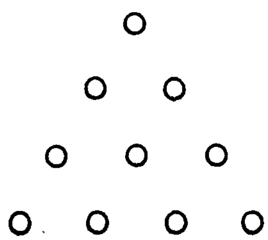
APPENDIX A (Continued)

Allowable moves: Move only one disc at a time; take only the top disc on a peg; never place a larger disc on top of a smaller one. (e.g., Karat, 1982; Levine, 1986)

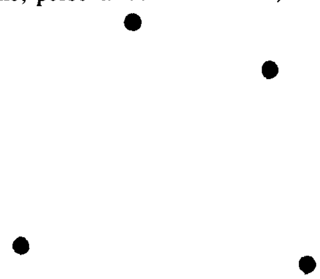
4. Three people play a game in which one person loses and two people win each round. The one who loses must double the amount of money that each of the other two players has at that time. The three players agree to play three games. At the end of the three games, each player has lost one game and each person has $8. What was the original stake of each player? (R. Thaler, personal communication, September 1986)
5. Next week I am going to have lunch with my friend, visit the new art gallery, go to the Social Security office, and have my teeth checked at the dentist. My friend cannot meet me on Wednesday; the Social Security office is closed weekends; the dentist has office hours only on Tuesday, Friday, and Saturday; the art gallery is closed Tuesday, Thursday, and weekends. What day can I do everything I have planned? (Sternberg & Davidson, 1982)

APPENDIX B
Insight Problems

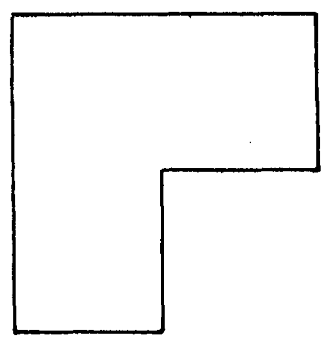
1. A prisoner was attempting to escape from a tower. He found in his cell a rope which was half long enough to permit him to reach the ground safely. He divided the rope in half and tied the two parts together and escaped. How could he have done this? (Experiments 1 and 2) (Restle & Davis, 1962)
2. Water lilies double in area every 24 hours. At the beginning of summer there is one water lily on the lake. It takes 60 days for the lake to become completely covered with water lilies. On which day is the lake half covered? (Experiments 1 and 2) (Sternberg & Davidson, 1982)
3. If you have black socks and brown socks in your drawer, mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure that you have a pair the same color? (Experiments 1 and 2) (Sternberg & Davidson, 1982)
4. The triangle shown below points to the top of the page. Show how you can move 3 circles to get the triangle to point to the bottom of the page. (Experiments 1 and 2) (deBono, 1969)



5. A landscape gardener is given instructions to plant 4 special trees so that each one is exactly the same distance from each of the others. How is he able to do it? (Experiments 1 and 2) (deBono, 1967)
6. A man bought a horse for $60 and sold it for $70. Then he bought it back for $80 and sold it for $90. How much did he make or lose in the horse trading business? (Experiment 2) (deBono, 1967)
7. A woman has 4 pieces of chain. Each piece is made up of 3 links. She wants to join the pieces into a single closed loop of chain. To open a link costs 2 cents and to close a link costs 3 cents. She only has 15 cents. How does she do it? (Experiment 2) (deBono, 1967)
8. Without lifting your pencil from the paper show how you could join all 4 dots with 2 straight lines. (Experiment 2) (M. Levine, personal communication, October 1985)



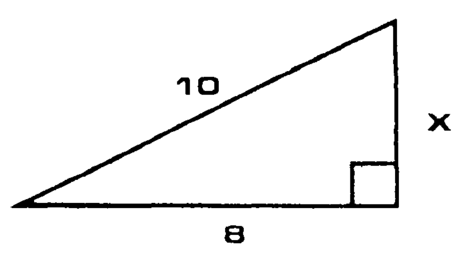
9. Show how you can divide this figure into 4 equal parts that are the same size and shape. (Experiment 2) (Fixx, 1972)



10. Describe how to cut a hole in a 3 × 5 in. card that is big enough for you to put your head through. (Experiment 2) (deBono, 1969)
11. Show how you can arrange 10 pennies so that you have 5 rows (lines) of 4 pennies in each row. (Experiment 2) (Fixx, 1972)
12. Describe how to put 27 animals in 4 pens in such a way that there is an odd number of animals in each pen. (Experiment 2) (L. Ross, personal communication, December 1985)

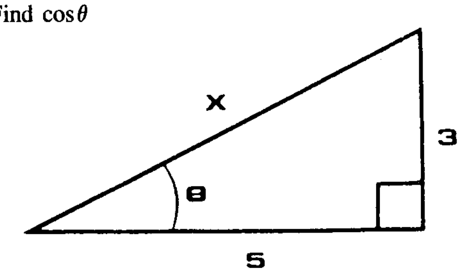
APPENDIX C
Math Problems (Taken from Travers, Dalton, Bruner, & Taylor, 1976)

1. $(3x^2+2x+10)(3x)$
2. $(2x+y)(3x-y) =$
3. Factor:
   $16y^2-40yz+25z^2$
4. Solve for x:



Appendix C (Continued)

1. $(3x^2+2x+10)(3x) =$
2. $(2x+y)(3x-y) =$
3. Factor:
   $16y^2-40yz+25z^2$
4. Solve for x:
5. $18x^2 + \frac{24x}{3x} =$
6. Factor:
   x^2+6x+9
7. Solve for x:
   $\frac{1}{5}x+10 = 25$
8. $\frac{-6x^2y^4}{3x^5y^3} =$
9. $\sqrt[3]{-27} =$
10. $\sqrt[6]{25} =$
11. Solve for x, y, and z:
    $x+2y-z = 13$
    $2x+y+z = 8$
    $3x-y = 2z = 1$
12. $\sqrt[3]{1000} =$
13. Solve for m:
    $\frac{m-3}{2m} - \frac{m-2}{2m+1} = 0$
14. $\sqrt{-121} =$
15. Solve for a and b:
    $3a+6b = 5$
    $2a-b=1$
16. $\sqrt[3]{16} =$
17. $(\sqrt[6]{2^2})(\sqrt[6]{2^3}) =$
18. $(a^2)(a^7) =$
19. $\frac{(a^2)}{(a^6)} =$
20. Find $\cos\theta$



(Manuscript received May 19, 1986;
revision accepted for publication September 25, 1986.)