A Theory of Problem Monitoring and Its Explanation of the Suddenness of Insight

Mark A. Changizi
Department of Computer Science
University College Cork—
National University of Ireland, Cork
Ireland
changizi@cs.ucc.ie

August 1, 1999

Abstract

The eureka, or aha!, phenomenon, associated with the suddenness of insight, is characterized by moments when the solution to a problem suddenly comes to the problem-solver without warning. Mathematical properties of problems related to a problem's susceptibility to solution via eureka moments are presented. An experiment carried out confirms three claims: (i) Eureka moments are not sufficient for insight, (ii) the eureka phenomenon is sometimes best explained by mathematical properties of problems, and (iii) other theories of the eureka phenomenon are not complete.

1 Introduction

This work's topic is not insight, but the suddenness of insight, which is sometimes called the eureka, or aha!, phenomenon. The eureka phenomenon is characterized by those occasions when the solution to a problem you are attempting to solve "hits" you without warning.

Insight and the eureka phenomenon are so intertwined that the phenomenon is sometimes taken to define insight [13, 10]. However, the phenomenon is arguably independent of insight ([27], p. 166) and I will argue that it is at least insufficient for insight. Metcalfe [12] carried out studies in which problem-solvers while working on problems give at regular intervals warmth-ratings representing how near they feel they are to the solution. In this paper I use Metcalfe's warmth-rating records by problem-solvers during problem-solving to operationally define the eureka phenomenon: Eureka moments are those moments when a problem-solver's warmth-rating jumps rapidly just before or as the problem-solver finds what he believes to be the solution. Problems are solved through non-eureka moments when the warmth-ratings rise incrementally during problem-solving. I present a theory of problem-solving and warmth-ratings which states that there are mathematical reasons explaining why for certain problems it is more difficult to give incremental warmth-ratings, and thus these problems are more susceptible to eureka moments. These mathematical observations motivate certain predictions concerning which problems are more likely to lead to a eureka moment. The experiment tests one of these predictions and provides confirmation of the theory. The experiment also confirms that eureka moments are insufficient for insight, since the problem in the experiment on which a eureka moment is experienced is not an insight problem by any standard. Finally, we will see that no other theory plausibly explains the results of the experiment, which implies that the conditions specified for eureka moments in every other theory are not necessary conditions for the phenomenon.

For the remainder of this section I present the psychological model, the model of what are problems, a formal notion of warmth-monitoring the solving of a problem, constraints on problem-solvers within the model attempting to warmth-monitor problems, and the theory for why problem-solvers experience eureka moments.

The Model

Problems are modeled as functions. For example, the function EVEN such that EVEN(x)=YES if x is even and NO otherwise is the problem of finding even numbers. While EVEN is the problem, EVEN(x) is an instance of the evenhood problem; e.g., EVEN(4) represents the question, "Is 4 even?"

Human problem-solvers are modeled as computational agents capable of computing only those functions—i.e., capable of solving only those problems—that are computable by a Turing machine. Human problem-solvers run "programs in the head," and in attempting to solve a problem they are assumed to run a program intended to compute the function corresponding to the problem. More accurately, human problem-solvers in the model attempt to solve problem-instances of problems, and they run the program intended for that problem on the problem-instance. The output of the program is the problem-solver's attempted answer to the problem-instance. Note that this model of human problem-solvers is just a weak computational model. Computational models are commonplace, and the only unique aspect of this one is that there are no specifics and so it is a generalization of every particular computational model. Also note that no claim has been made concerning what the program in the head for the function is computing. A single program used to solve an instance of a general problem does not imply that a single sort of problem-solving technique is employed, or that a search through just one problem search space (or problem representation) is carried out. Although at some level of explanation it might be useful to, if possible, tease apart the single program into multiple programs used in different circumstances, there is under this model ultimately just a single (possibly very complicated) program being run. Finally, there is no commitment in this model of problem-solving that the problem-solver is conscious of the steps being carried out.

Warmth-monitorability

Motivated by the eureka phenomenon I have developed a theory of "warmth-monitorable functions" ([3], Sections 1 and 3) where, informally, a function is *warmth-monitorable* if there exists a program computing it that is capable of keeping tabs on how near it is to halting. Rather than giving the formal definition of warmth-monitorability, which is beyond the scope of this paper, I communicate the idea with several examples.

Suppose you are asked to compute the function $f(x) = x + 7$ and that each step of your computation may only consist in adding one to a number. No matter what input, or problem-instance, I give you, all you need to do is take seven successors of the input and you are finished. You can be sure that it will take you just seven steps to get to your solution, and you will not be at all surprised when you get to it. You are able to "keep tabs" on your progress: "Seven steps left. Six steps left. Five steps left. Etc." And at each stage you have a quantitative measure of how "near" you are to finishing. Imagine that you have a "warmth clock" in your head that is set to a fixed value before seeing the input, and is decremented so that when the clock hits zero the solution is found. You know how near, or warm, you are to finishing the problem-solving by noting how near the clock is to zero. In the case of $f(x) = x + 7$ the warmth clock may be initially set to seven, and this suffices for every instance of the problem. I say that this function has a warmth-monitoring *degree* of seven.

Now consider a more difficult example: $g(x) = 2x$. Again assume that each step in computing the function may only consist in adding one to a number. No matter what input x I give you, it will take x applications of successor to get to the solution. Again you will not be surprised when the solution is reached, and this is because you are able to count down, "x steps left. $x - 1$ steps left. Etc." What should the warmth clock be initialized to, however? If we try to set it to any particular natural number, there will be inputs x larger than that number and the warmth clock will then get to zero before the function is computed. This is where transfinite ordinals enter (do not worry if you do not understand ordinals—the informal presentation should communicate the main idea without the reader needing to understand ordinals). ω is the least ordinal greater than every natural number, and it suffices to initialize the warmth clock to ω for function g. On any input x, your first act is to decrement the clock from ω to x. Then it takes x computations to finish. The warmth-monitoring degree in this case is, then, ω. While seven computations are required to solve f, one *decision* about how many more computations to engage in is required to solve g: "I will allow myself only one decision concerning how many computation steps I will allow myself, and I must make the decision immediately." For f, on the other hand, the sentence is, "I will allow myself exactly seven computation steps." The ordinal ω indicates that one decision is needed concerning how many steps are to be taken in solving the problem.

Let us move to a still more difficult warmth-monitorable function: $h(x, y, z) = x + y + z$. To solve this problem it suffices to take y successors of z, getting $y + z$, and then to take x successors of $y + z$.

There are no surprises here either, since you realize that you must first count down through y, and then count down through x; then and only then will you be finished. Although this is still warmth-monitorable, there is a sense in which it seems a bit more difficult to warmth-monitor than f or g. What should the warmth clock be initialized to? To help guide us in its initialization realize that you need to make two decisions in order to solve h: the first is to allow yourself y steps for taking the successors of z, and the second is to allow yourself x steps for taking the successors of $y + z$. You cannot just make one decision to, say, allow yourself $x + y$ steps for that many successors of z, since you have not yet computed the value of $x + y$. Therefore, it is not sufficient to initialize the warmth clock to ω because this allows only one decision. $\omega + \omega$ (which is the least ordinal greater than every ordinal of the form $\omega + n$) might seem to be the proper initialization of the warmth clock, and indeed it is. On any three inputs x, y, z your first act is to decrement the clock from $\omega + \omega$ to $\omega + y$. After y successors of z are taken the clock is at ω. You then decrement the clock from ω to x and take x successors of $y + z$, at which point the clock is at zero and the problem has been solved. "I will allow myself two decisions about how many computations I will allow myself, the first which must be made immediately." The warmth-monitoring degree of h is $\omega + \omega$ (or $\omega \cdot 2$), and this indicates that two decisions are required.

I conclude this series of examples with just one more. Take $t(x) = x^2$. To compute this using only the successor it suffices to take x successors of x to receive $2x$, then x successors of $2x$ to receive $3x$, then x successors of $3x$ to receive $4x$, and so on x times. Again there is no surprise. Let us look at this at the level of decisions. Your first decision needs to be about how many times you will take x successors of things; in fact, you will need to do this x times. Taking x successors requires a decision, however. Thus, your first decision is about how many decisions you require: "I will allow myself exactly one (second-order) decision concerning how many (first-order) decisions I will allow myself." It turns out that the warmth-monitoring degree for this is ω^2 (which is the least ordinal greater than ordinals of the form $\omega \cdot n$). On input x, your first decision is that you need x decisions, so you decrement the clock to $\omega \cdot x$. You then use up your first of these x decisions by deciding to allow yourself x successors of x, so you decrement the warmth clock to $\omega \cdot (x - 1) + x$. Once you have taken x

successors of x the clock is at $\omega \cdot (x-1)$ and you have computed $2x$. You then use up your second decision by deciding to allow yourself x successors of $2x$. You do this by decrementing the clock to $\omega \cdot (x-2)+x$ and taking x successors of $2x$, at which point the clock is at $\omega \cdot (x-2)$. Etc.

One may wonder why the warmth clock must be initialized to a fixed value for every input. Why not allow the clock to be set *after* seeing the input? For example, for $t(x) = x^2$ it would then suffice to, upon seeing input x, set the clock to $\omega \cdot x$. Suppose that you decide that this is how you would like to engage in your warmth-monitoring, and that α_x is the ordinal you set the clock to upon seeing input x. Now let β be the least ordinal greater than each α_x. It is then no restriction for you to, instead, initialize your warmth clock to β, and then proceed as you desired and decrement the clock to α_x upon seeing input x. Requiring that the clock be initialized before seeing the data helps to communicate, however, the fact that it must be possible for a *single* problem-solver to properly set the clock for *every* input. It is not good enough to have a function f such that for each input x there exists a program that can compute $f(x)$ as well as warmth-monitor it.

So long as a function has a warmth-monitoring degree, it is possible for there to be absolutely no surprise when computing it. I say "possible" because even for $f(x) = x+7$ one *would* be surprised if one wrongly thought it should take one thousand steps. And although I have only shown examples with warmth-monitoring degree as high as ω^2, the hierarchy continues long past this: ω^ω, ω^{ω^ω}, etc., are part of the hierarchy. What makes such functions warmth-monitorable is that at each stage of the computation the warmth clock records how much more thinking and decision-making is required, which is quantitatively measured in terms of ordinals.

I do not claim that warmth-monitorability is an accurate descriptive model of the way human problem-solvers actually acquire their warmth-ratings when they are problem-solving. Rather, this notion of warmth-monitorability motivates the idea that some problems are intrinsically more difficult to accurately acquire incremental warmth-ratings records, and that there is a positive correlation between the warmth-monitoring degree of a problem and its difficulty for humans to give incremental warmth-rating records.

I have implicitly been assuming that the problem-solver correctly acquires programs computing the functions he intends to compute.

This hides at least two idealizations. One, the problem-solver might not acquire the correct program; the program the problem-solver chooses might compute some different function. Even so, such a program may be useful to the problem-solver for solving the original function (the original problem) at least in some cases (i.e., it may be a heuristic). Two, given a problem-instance it is not always clear what is the problem for which it is an instance? For example, given the problem-instance of the coin problem (A coin collector is presented with a bronze coin on which is written "450 B.C.". He knows immediately that the coin is a fake. Why?), what is the general problem? It is entirely unclear. Fortunately it is not essential in the model to know what the general problem actually is. The model presumes that the problem-solver is running a program in the head on the input that is the instance of the problem. The function that really matters is whatever function it is that that program is computing. It is ultimately the warmth-monitorability of this function that is relevant, although it is convenient to assume that this function is the same as the function an instance of which the problem-solver is attempting to solve.

The Theory

Why might a problem-solver fail to give incremental warmth-ratings records? Why might the solution time come as a surprise to a problem-solver? That is, why might there be eureka moments? There are at least three reasons related to the intrinsic mathematical nature of problems and their warmth-monitorability.

Impossible to monitor. First, the function the problem-solver is computing might be of the mathematical nature that it is not possible, even in principle, to have incremental warmth-rating records; the time of solution might be guaranteed to be a surprise. This possibility is motivated by the fact that it is a possibility for warmth-monitorability: some computable functions are not warmth-monitorable at all. This is a consequence of Theorem 9 in Changizi [3] which says that there are computable functions outside the hierarchy of warmth-monitorable functions. There are artificial extensions of my hierarchy that can be concocted, but a result due to Feferman [6], Theorem 5.4) prevents there being an extension that violates the truth of my claim. Some problems, then, *require* a eureka moment to solve them, and this is true whether the problem-solver is human or HAL from *2001: Space*

Odyssey.

Difficult to monitor. Second, it may be possible to solve the problem with incremental warmth-rating records, thereby not experiencing a eureka moment, but it may be too taxing to do so. This possibility is motivated by the fact that for warmth-monitorability there are functions with higher and higher warmth-monitoring degrees. At some point in the hierarchy, human problem-solvers are likely to find it too complicated to maintain warmth-monitoring—for example, for problems with warmth-monitoring degree quantified by a tower of one billion ω's.

Tricky to monitor. Third and last, it may be possible to solve the problem with incremental warmth-rating records, and it may be within the problem-solver's ability to do so, but the problem-solver may not know how to do so nevertheless because how to do so may be far from obvious. This possibility is motivated by the fact that for warmth-monitorability there are functions that are quite easily warmth-monitorable, but it is not obvious that one can do so. Warmth-monitorability does not entail that it is a simple matter to determine the warmth-monitoring degree or the method by which to decrement the ordinals. Similarly, the mere fact that it is possible to easily solve a problem and report incremental warmth-ratings does not mean it is easy to know how to do so.

My theory of the eureka phenomenon is this: Eureka moments—operationally defined as the rapid rise in warmth-rating reports in problem-solving just before or as the problem-solver arrives at what he believes is the solution—are explained by the intrinsic mathematics underlying problems, and three possible related reasons for eureka moments on a problem are that the problem is (a) impossible to monitor, (b) difficult to monitor, and (c) tricky to monitor, as discussed above.

The three parts of the theory—(a), (b) and (c)—each make certain predictions. (a) predicts that human problem-solvers will experience eureka moments on problems that are impossible to warmth-monitor. This is not easy to test since the experimenter must concoct problems that are both impossible to warmth-monitor yet simple enough for subjects to solve. I have not found good candidate problems for this yet. (b) predicts that for problems at some sufficiently high level of warmth-monitoring degree, problem-solvers will be more likely to experience eureka moments. I have found this difficult to test as well.

It might be that as long as the problem-solver knows how to monitor it, he will do so provided it is below a certain threshold of difficulty. If this were so, then experimenting would show no effect until the threshold is reached. Furthermore, the threshold may be so high that finding suitable problems for subjects is difficult. In this paper I do not test (b) either. (c) predicts that problems that are easy but tricky to warmth-monitor are likely to give problem-solvers eureka moments. This is the most susceptible prediction to empirical test, and the experiment aims to confirm this.

Successful confirmation via the experiment of this aspect of the theory would only demonstrate that the theory is an explanation for some cases of eureka moments. Since we will see that no other theory explains these cases of eureka moments, successful confirmation of my theory in this regard would argue that my theory is the *best* explanation for some cases of eureka moments. In this paper I present no data suggesting that my theory is able to explain all cases of eureka moments; my theory may provide sufficient but possibly not necessary conditions for eureka moments. This is a matter for future study.

With (c) above in mind, I invented two problems in arithmetic: a not-tricky-to-monitor problem and a tricky-to-monitor problem. Except for this difference the problems are similar in kind, in particular their respective functions are similar in warmth-monitorability. The not-tricky-to-monitor problem-instance presented to subjects is as follows:

> $x = 1022$. $x := \frac{x}{2}$ (i.e., x is halved), and then $x := x - 1$ (i.e., x is decremented by one). If you continue this (halving and decrementing) there is a point at which x becomes less than or equal to 0. When this first happens is $x < 0$ or $x = 1$? Solve the problem by beginning with the first number and modifying it as instructed until you find the answer.

The correct solution process goes as follows:

- $\frac{1022}{2} = 511$.
- $511 - 1 = 510$.
- $\frac{510}{2} = 255$.
- $255 - 1 = 254$.
- $\frac{254}{2} = 127$.
- $127 - 1 = 126$.

- $\frac{126}{2} = 63$.
- $63 - 1 = 62$.
- $\frac{62}{2} = 31$.
- $31 - 1 = 30$.
- $\frac{30}{2} = 15$.
- $15 - 1 = 14$.
- $\frac{14}{2} = 7$.
- $7 - 1 = 6$.
- $\frac{6}{2} = 3$.
- $3 - 1 = 2$.
- $\frac{2}{2} = 1$.
- $1 - 1 = 0$.
- Answer is 0.

This is not tricky to monitor since although the problem-solver may not know along the way what the answer will be, he knows that he is nearing the solution since the numbers are inexorably getting nearer to zero at which point he will learn whether the first such point is < 0 or $= 0$. There is no surprise and the problem-solver is expected to record incremental warmth-ratings.

The tricky-to-monitor problem-instance presented to subjects is as follows:

> $x = \frac{30}{32}$. $x := 2x$ (i.e., x is doubled), and then $x := x-1$ (i.e., x is decremented by one). If you continue this (doubling and decrementing), there is a point at which x becomes less than or equal to 0. When this first happens is $x < 0$ or $x = 1$? Solve the problem by beginning with the first number and modifying it as instructed until you find the answer.

The correct solution process goes as follows:

- $\frac{30}{32} \cdot 2 = \frac{60}{32}$.
- $\frac{60}{32} - \frac{32}{32} = \frac{28}{32}$.
- $\frac{28}{32} \cdot 2 = \frac{56}{32}$.
- $\frac{56}{32} - \frac{32}{32} = \frac{24}{32}$.
- $\frac{24}{32} \cdot 2 = \frac{48}{32}$.

- $\frac{48}{32} - \frac{32}{32} = \frac{16}{32}$.
- $\frac{16}{32} \cdot 2 = \frac{32}{32}$.
- $\frac{32}{32} - \frac{32}{32} = 0$.
- Answer is 0.

This is tricky to monitor since it is difficult for the problem-solver to tell what is going on. The numbers are getting larger and smaller, and then "out of the blue" it jumps all the way down to 0. There are fewer steps required for this problem than the previous because I found that subjects take longer to solve the individual steps of this problem than in the other.

2 Experiment

Method

Materials: Subjects were given two sheets of paper stapled together on which were written the not-tricky-to-monitor and tricky-to-monitor problem-instances. Half of the subjects received papers with the not-tricky-to-monitor problem-instance on the first page, and the other half of the subjects received papers with the tricky-to-monitor problem-instance on the first page. On each piece of paper were 56 boxes (in a 7 by 8 grid) large enough to be labeled with numbers from 1 through 10.

Subjects: The subjects were 88 students of computer science asked to participate in a problem-solving experiment.

Procedure: The subjects were instructed similar to as in Experiment 1 of Metcalfe [12]. They were told to begin on the problem on the first page given to them, moving straight to the second problem once finished. They were also told that every 15 seconds they would be asked to write down a number from 1 through 10 in a box representing the degree to which they feel they are close to the solution (a warmth-rating), where 1 means they have no idea what the solution is and 10 means they know what the solution is. They were instructed that if they ever wrote a 10 then that means they have completed the problem. Subjects were also instructed to show their work and circle their answer.

Results

I say that a subject's data is "valid" if (i) the subject's warmth-rating protocol exists, is interpretable, and has at least four ratings before the '10', (ii) the subject's answer and work are shown, and (iii) the subject's work indicates that the subject's answer was acquired via the "automatic" manner as directed (rather than an analytical solution). Note that a subject may have arrived at an incorrect solution yet the subject's data be deemed valid. This occurs when the subject's error does not change the degree to which monitoring the problem-solving is tricky. For example, if a subject divides 1022 by 2 and incorrectly gets 510, but continues step by step until < 0, then there is, from the subject's point of view, just as much or little surprise at when the solution comes.

Probability level $p < .05$ will be taken as the criterion for statistical significance in this paper. No statistically significant differences were found concerning the order in which the problems were presented to the subjects, and I will not mention this distinction any further.

There were 65 subjects with valid data for the tricky-to-monitor problem and 71 subjects with valid data for the not-tricky-to-monitor problem. 59 and 56 of these were correct for, respectively, the tricky- and not-tricky-to-monitor problems. The following was computed for each set of data (valid and correct-valid): For each problem the warmth-rating 15 seconds before the '10' was recorded. A higher warmth-rating at 15 seconds before the '10' indicates a more incremental warmth-rating protocol.

For each warmth-rating value of n (1 through 9) the relative frequency that n occurred in a not-tricky-to-monitor problem was computed from the data. The correlation was calculated comparing these relative frequencies to the warmth-rating values. A positive correlation means that higher warmth-rating (and thus more incremental warmth-rating protocols) are more likely to be for the not-tricky-to-monitor problem, and is what my theory predicts. The correlation among valid data was .68 ($t = 2.46$, $p < .05$ on a two-tailed test), and among correct-valid data was .60 ($t = 2.00$, $p < .05$ on a one-tailed test).

Among the 55 subjects who validly answered both problems, for the warmth-rating 15 seconds before the '10' there were 28 subjects with a smaller warmth-rating for the tricky-to-monitor problem, 9 subjects with a greater warmth-rating for the tricky-to-monitor problem, and 18 subjects with the same warmth-rating. Under a sign test this trend is significant ($t = 2.63$, $p < .005$ on a one-tailed test). Of the 43 of these subjects who correctly answered both problems, the three values are 17, 9, and 18, which is not significant $< .05$ ($t = 1.37$, $p \approx .086$ on a one-tailed test) despite still trending as predicted.

Discussion

A eureka moment is not sufficient for insight. The tricky-to-monitor problem-instance which leads to a eureka moment is by no means an insight problem-instance (except where one *defines* insight as coextensive with eureka moments). To solve it one need only continue to double and decrement without thought, waiting until the result is less than or equal to zero. Our pretheoretic notion of insight, however slippery and vague, simply does not apply to this tricky-to-monitor problem-instance.

Other theories are not complete. Many theories attempting to explain the eureka phenomenon postulate that a sort of leap occurs, e.g., representation change [4, 7, 10, 15, 16], perceptual gestalt ground reversals [5], and perceptual gap filling [8]. Other theories presume that some sort of fixation is overcome [1, 5, 21, 22, 31]. The tricky-to-monitor problem-instance is solved by no leap, as can be verified by looking at the subjects' work: they arrive at the solution in the incremental fashion I gave earlier. There is also clearly no fixation involved. Non-incremental warmth-ratings do not, then, imply non-incremental problem-solving, which is one of the main theses of Weisberg [28, 29, 30, 24, 25, 26, 27] (see also Perkins [18]). Others explain the eureka phenomenon through subconscious or non-reportable mental processes [2, 17, 19, 32]; see Schooler et. al. ([19], p. 168) for references to early such theories. Non-reportable mental processes do not seem plausibly relevant for the tricky-to-monitor problem-instance. Still other directions include the notion that the eureka phenomenon is related to how unexpected the solution is [14], that it is connected with a sort of nondeterminism in thinking [9], and that it is triggered in part by external cues [11, 20, 23]. Again, none of these explanations seem at all applicable to the tricky-to-monitor problem-instance. Each of these theories therefore specifies conditions which, even if sufficient for eureka moments, are not necessary for eureka moments; the theories are incomplete.

My theory is sufficient. My theory explains why there is a eureka moment in the tricky-to-monitor problem-instance and why there is not a eureka moment in the not-tricky-to-monitor problem-instance via part (c) of the theory: problems for which it is not obvious how to warmth-monitor are more likely not to acquire incremental warmth-rating records. The experiment is therefore a confirmation of my theory.

3 Conclusion

I have now demonstrated my three central claims: eureka moments are not sufficient signs of insight, prior theories of the eureka phenomenon are incomplete, and my theory claiming that eureka moments are caused, at least sometimes, by the underlying mathematics of problems does explain the experiment and thereby receives some measure of confirmation. At the level of speculation, perhaps the strong link between eureka moments and insight problems is partly explained by the fact that human problem-solvers successful at insight problems sometimes implement programs in the head that carry out computations that are either impossible, difficult, or tricky to monitor.

It should be stated once again that I have operationally *defined* the eureka phenomenon as those moments when warmth-ratings suddenly increase just before or as the problem-solver arrives at what he believes is the solution. One may very plausibly complain that there are affective features of true eureka moments that such an operational definition does not capture, and that I have therefore not provided a theory giving sufficient criteria for true eureka moments. This seems to me a justifiable worry, and my conclusions should be considered in its light. However, if the eureka phenomenon consists of more than that of my operational definition of it, then, as far as I know, no theory provides conditions sufficient to explain true eureka moments.

References

[1] J. R. Anderson. *Cognitive Skills and Their Acquisition.* Erlbaum, 1981.

[2] K. S. Bowers, G. Regehr, C. Balthazard, and K. Parker. Intuition in the context of discovery. *Cognitive Psychology*, 22:72–110, 1990.

[3] M. A. Changizi. Self-monitoring machines and an ω^ω-hierarchy of loops. *Information and Computation*, 128(2):127–138, 1996.

[4] J. E. Davidson. The suddenness of insight. In R. J. Sternberg, editor, *The Nature of Insight*, pages 125–156. MIT Press, 1995.

[5] P. Ellen. Direction, past experience, and hints in creative problem solving: Reply to Weisberg and Alba. *Journal of Experimental Psychology: General*, 111(3):316–325, 1982.

[6] S. Feferman. Classifications of recursive functions by means of hierarchies. *Trans. Amer. Math. Soc.*, 104:101–122, 1962.

[7] M. L. Gick and R. S. Lockhart. Cognitive and affective components of insight. In R. J. Sternberg, editor, *The Nature of Insight*, pages 197–228. MIT Press, 1995.

[8] H. E. Gruber. Insight and affect in the history of science. In R. J. Sternberg, editor, *The Nature of Insight*, pages 397–432. MIT Press, 1995.

[9] P. N. Johnson-Laird. Freedom and constraint in creativity. In R. J. Sternberg, editor, *The Nature of Creativity: Contemporary Psychological Perspectives*, pages 202–219. Cambridge University Press, 1988.

[10] C. A. Kaplan and H. A. Simon. In search of insight. *Cognitive Psychology*, 22:374–419, 1990.

[11] P. Langley and R. Jones. A computational model of scientific insight. In R. J. Sternberg, editor, *The Nature of Creativity: Contemporary Psychological Perspectives*, pages 176–201. Cambridge University Press, 1988.

[12] J. Metcalfe. Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12(4):623–634, 1986.

[13] J. Metcalfe. Intuition in insight and noninsight problem solving. *Memory and Cognition*, 15(3):238–246, 1987.

[14] N. Mori. A factor related to the strength of an "aha!" experience. *Japanese-Psychological Research*, 38(2):85–89, 1996.

[15] S. Ohlsson. Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, 25:117–129, 1984.

[16] S. Ohlsson. Restructuring revisited: Summary and critique of the gestalt theory of problem solving. *Scandinavian Journal of Psychology*, 25:65–78, 1984.

[17] S. Ohlsson. Information-processing explanations of insight and related phenomena. In M. Keane and K. Gilhooley, editors, *Advances in the Psychology of Thinking*, volume 1, pages 1–44. Harvester-Wheatsheaf, 1992.

[18] D. N. Perkins. *The Mind's Best Work*. Harvard University Press, 1981.

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

[20] C. M. Seifert, D. E. Meyer, N. Davidson, A. L. Patalano, and I. Yaniv. Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg, editor, *The Nature of Insight*, pages 65–124. MIT Press, 1995.

[21] H. A. Simon. Scientific discovery and the psychology of problem solving. In R. Colodny, editor, *Mind and Cosmos*, pages 22–40. University of Pittsburgh Press, Pittsburgh, 1966.

[22] S. M. Smith. Getting into and out of mental ruts: A theory of fixation, incubation, and insight. In R. J. Sternberg, editor, *The Nature of Insight*, pages 229–252. MIT Press, 1995.

[23] G. Wallas. *The Art of Thought*. Harcourt Brace Jovanovich, 1926.

[24] R. W. Weisberg. *Creativity, Genius and Other Myths*. Freeman, 1986.

[25] R. W. Weisberg. Problem solving and creativity. In R. J. Sternberg, editor, *The Nature of Creativity: Contemporary Psychological Perspectives*, pages 148–176. Cambridge University Press, 1988.

[26] R. W. Weisberg. Metacognition and insight during problem solving: Comment on Metcalfe. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 18(2):426–431, 1992.

[27] R. W. Weisberg. Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg, editor, *The Nature of Insight*, pages 157–196. MIT Press, 1995.

[28] R. W. Weisberg and J. W. Alba. An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General*, 110(2):169–192, 1981.

[29] R. W. Weisberg and J. W. Alba. Gestalt theory, insight, and past experience: Reply to Dominowski. *Journal of Experimental Psychology: General*, 110(2):193–198, 1981.

[30] R. W. Weisberg and J. W. Alba. Problem solving is not like perception: More on gestalt theory. *Journal of Experimental Psychology: General*, 111(3):326–330, 1982.

[31] R. S. Woodworth. *Experimental Psychology.* Henry Holt, 1938.

[32] I. Yaniv and D. E. Meyer. Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13:187–205, 1987.