THINKING & REASONING
2021, VOL. 27, NO. 1, 1–37
https://doi.org/10.1080/13546783.2019.1705912

Routledge
Taylor & Francis Group

The temporal dynamics of insight problem solving – restructuring might not always be sudden

Merim Bilalić[a], Mario Graf[b], Nemanja Vaci[c] and Amory H. Danek[d]

[a]Department of Psychology, University of Northumbria at Newcastle, Newcastle upon Tyne, UK; [b]Institute of Psychology, University of Klagenfurt, Klagenfurt, Austria; [c]Department of Psychiatry, University of Oxford, Oxford, UK; [d]Department of Psychology, University of Heidelberg, Heidelberg, Germany

ABSTRACT

Insight problems are likely to trigger an initial, inappropriate mental representation, which needs to be restructured in order to find the solution. Despite the widespread theoretical assumption that this restructuring process happens suddenly, which leads to the typical Aha! experience, the evidence is inconclusive. Among the reasons for this lack of clarity is a reluctance to measure solvers' subjective experience of the solution process. Here, we overcome previous methodological problems by measuring the dynamics of the solution process using eye movements in combination with the subjective Aha! experience. Our results demonstrate that in a problem that requires restructuring of the initial mental representation, paying progressively more attention to the crucial elements of the problem often preceded the finding of the solution. Most importantly, the sooner solvers started paying attention to the crucial elements, the less sudden and surprising the solution felt to them. The close link between the eye movement patterns and self-reported Aha! experience in the present study underlines the necessity of measuring both the cognitive and the affective components of insight to capture the essence of this phenomenon.

ARTICLE HISTORY Received 16 May 2019; Accepted 10 December 2019

KEYWORDS Insight; Aha! experience; problem solving; restructuring; eye movements

Introduction

The temporal dynamics of problem solving processes have been an important topic in cognitive science. Most problems are assumed to be solved gradually, by putting together pieces of information in order to arrive at the solution (Newell & Simon, 1972). "Insight problems" are believed to

CONTACT Merim Bilalić merim.bilalic@northumbria.ac.uk

represent a special type of problem because, in these cases, the solution often comes seemingly out of nowhere (Duncker, 1945), although the problem appeared unsolvable just a moment ago. To be solved, insight problems are thought to require a fundamental, sudden change in the way the problem is perceived, a process called restructuring (Ohlsson, 1992; Wertheimer, 1925). This should lead to a non-monotonic, discontinuous problem solving process (Danek, 2018) and is often accompanied by an Aha! experience, the phenomenological component of insight problem solving. However, not only does the evidence for the suddenness of the restructuring process remain controversial, but the connection between the subjective Aha! experience and the dynamics of problem solving is unknown. Here, we demonstrate that people pay increasingly more attention to the crucial elements of a problem that requires restructuring long before they actually find the solution. We do, however, also show that solvers who started paying attention to the crucial aspects of the problem earlier also report less of the Aha! experience compared to solvers who paid attention to the crucial aspect just before they found the solution.

According to Gestalt psychologists, the switch between the initial, incorrect problem representation and the correct one is unexpected and leaves an unmistakable phenomenological trace in solvers' experience: "The decisive points in thought-processes, the moment of sudden comprehension, of the 'Aha!,' of the new, are always at the same time moments in which such a *sudden restructuring* [emphasis added] of the thought-material takes place" (Duncker, 1945, p. 29). The restructuring from the initial, incorrect mental representation to the correct one is the key component in modern theories, such as the Representational Change Theory, or RCT (Knoblich, Ohlsson, Haider, & Rhenius, 1999; Ohlsson, 1984, 1992, 2011). It is also fairly well established that the restructuring process, leading to the realization of the solution, often elicits the Aha! experience (Kaplan & Simon, 1990).

Classical theories of insight assume that restructuring occurs abruptly (Davidson, 1995; Duncker, 1926), which is also the main reason why insight problems are viewed as being qualitatively different from ordinary problems where solutions are obtained gradually. For example, Metcalfe's series of studies provided the first evidence for this claim by showing that solvers cannot predict their solution progress (measured as feelings-of-warmth ratings) on a set of insight problems, in contrast to algebra problems (Metcalfe, 1986; Metcalfe & Wiebe, 1987). Similarly, a recent study found that feelings-of-warmth increased more abruptly for solutions to compound remote associates (CRAs) for which an Aha! had been self-reported than for the solutions where this feeling was missing (Kizilirmak et al., 2018). Another study by Smith and Kounios (1996) provided evidence for "all-or-none processing" (p. 1443) in the solving of anagrams because solvers were not aware of any partial solution information before the moment of the actual solution, as opposed to a more incremental process where partial solution information should be available.

Although the sudden nature of the underlying restructuring process is one of the main theoretical assumptions about insight, the evidence for this claim is inconclusive. Ohlsson (1992) even hypothesized that "the sudden appearance of the complete solution in consciousness is an illusion caused by our lack of introspective access to our cognitive processes (…)" (p. 17). To truly answer the question about the temporal nature of insight, both the cognitive component of insight (restructuring) and the affective component (Aha! experience) must be measured. An objectively sudden change in the problem representation would be expected to co-occur with subjectively perceived suddenness of the solution process or an Aha! experience. There is a wealth of studies investigating Aha! experiences without directly assessing the cognitive component (e.g., Jung-Beeman et al., 2004; Kizilirmak, Galvao Gomes da Silva, Imamoglu, & Richardson-Klavehn, 2016; Rothmaler, Nigbur, & Ivanova, 2017; Webb, Little, & Cropper, 2018). Several other studies focused on restructuring, but typically without measuring the dynamics of the solving process or the Aha! experience (e.g., Ash, Jee, & Wiley, 2012; Ash & Wiley, 2006, 2008; Fleck & Weisberg, 2013; MacGregor & Cunningham, 2009). There are only a few studies that directly measured both the cognitive and the affective components of insight (Cushen & Wiley, 2012; Danek, Williams, & Wiley, 2018; Ellis, Glaholt, & Reingold, 2011).

The cognitive component of insight: restructuring the mental problem representation

Obviously, fleeting changes in solvers' mental problem representation are difficult to assess. Previous studies that attempted to measure the temporal dynamics of restructuring used different types of trace data. Some used repeated ratings of problem elements, either with regard to their similarity (Durso, Rea, & Dayton, 1994) or with regard to their relevance for the solution (Cushen & Wiley, 2012; Danek et al., 2018). Others employed eye movement recordings (Ellis et al., 2011; Knoblich, Ohlsson, & Raney, 2001; Tseng, Chen, Chen, Sung, & Chang, 2014) or solvability judgments (Novick & Sherman, 2003). In some of these studies, both incremental and sudden solution patterns were found (Cushen & Wiley, 2012; Danek et al., 2018; Novick & Sherman, 2003), whereas other studies found only incremental patterns (Durso et al., 1994; Ellis et al., 2011).

In the first attempt at measuring the temporal dynamics of the restructuring process, Durso et al. (1994) asked participants to rate the relatedness of word pairs in a word puzzle during the problem solving process. On average, solution-relevant pairs were rated as progressively more similar as participants approached solution, from which the authors concluded that "[l]ike dynamite, the insightful solution explodes on the solver's cognitive landscape with breathtaking suddenness, but if one looks closely, a long fuse warns of the impending reorganization." (Durso et al., 1994, p. 98). Similar evidence was provided by Novick and Sherman (2003; Experiment 2), who asked participants to indicate within a short time window (250 ms after stimulus offset) whether presented anagrams were solvable. Although they could not find the solution in such a short time, participants were increasingly better at differentiating between solvable and unsolvable anagrams as the presentation time of the anagrams increased. In accordance with Durso et al., the authors concluded that solvers gradually accumulated information relevant for solving the anagrams.

Unfortunately, both studies suffer from methodological problems. Besides the lack of a subjective measure of the phenomenology, the Durso et al. study looked only at averaged group data. As pointed out by Cushen and Wiley (2012), analyzing the data on an aggregate level could have averaged out sudden patterns present in individuals (for a similar phenomenon in skill acquisition, see Heathcote, Brown, & Mewhort, 2000). Another difficulty is related to the repeated rating procedure in the Durso et al. study, which may have drawn attention to important elements and served as cues. Novick and Sherman did not measure the actual finding of a solution because solutions were provided directly after the solvability judgments. Furthermore, a similar anagram study could not replicate the gradual accumulation of information in anagrams (Smith & Kounios, 1996).

Following up on the Durso et al. study, Cushen and Wiley (2012) asked participants to rate individual problem elements (the circles of the Triangle of Circles problem) with regard to how important for solution they were. In contrast to Durso et al., the data was analyzed on an individual level instead of aggregating across all solvers' ratings, thereby avoiding the possibility of obscuring patterns only present in individuals. The patterns of changes in the relevant elements were rated as incremental or "insight-like", and both sudden (68% of all cases) and incremental (32%) patterns were found. Similarly, a recent study measured participants' problem representations while they solved a set of 18 magic tricks (Danek et al., 2018). Across three repeated viewings of each magic video clip, action verbs (including one that implied the solution) had to be rated with regard to how important for solution they were. 14% of all correct solutions were categorized as showing a sudden pattern and 10% as showing an incremental pattern, with the majority of solutions showing other patterns (e.g., flat, decreasing or zigzag).

As is obvious from this literature review, no definite conclusions about the temporal dynamics of the cognitive restructuring process underlying

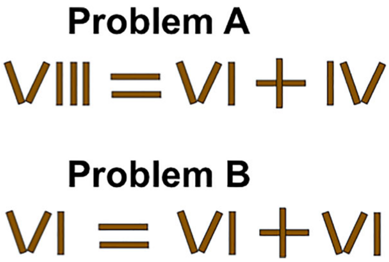


Figure 1. Matchstick arithmetic problems used.

insight can be drawn from these studies. An alternative, perhaps less intrusive method than ratings may be the recording of eye movements. Eye movements provide an objective measure of cognitive processes, as they are strongly linked to attention (e.g., Just & Carpenter, 1976; Rayner, 1995; Reingold, Charness, Schultetus, & Stampe, 2001). Eye fixations reveal whether people paid attention to certain features of the problem, and if so for how long, even when they might not remember or even concurrently report that they are paying attention to these elements (Bilalić & McLeod, 2014; Bilalić, McLeod, & Gobet, 2008, 2010; Kuhn & Land, 2006; Kuhn, Tatler, & Cole, 2009). This is particularly relevant in the present situation, where it is possible that people are not aware of the dynamics of their solution process.

The seminal study by Knoblich, Ohlsson, and Raney (2001) demonstrated the usability of eye movement recordings in insight research and is an important precedent for our study. We used two problems from the matchstick arithmetic domain which belong to the same problem type (standard type, ST, and constraint relaxation type, CR3) and are structurally identical to problem A (IV = III + III) and B (III = III + III) from Knoblich et al.'s study, the only difference being different values (see Figure 1). Knoblich et al. compared these two problem types because standard type problems are not supposed to require any restructuring, whereas constraint relaxation type problems do. They found that, for Problem B (constraint relaxation type), both solvers and nonsolvers began by examining the values and spent most of their time doing so. This can be taken as an indication that participants were using an initial incorrect problem representation where only values can be changed. Only in the final third of the problem solving period did later solvers change their mental representation, as demonstrated by their spending more time on the operators and less on values. In contrast, nonsolvers remained stuck and fixated in their initial mental representation as they continued to examine the values. Another eye tracking study found similar results for the same Problem B (Tseng, Chen, Chen, Sung, & Chang, 2014), but did not include a comparison with the second matchstick problem.

Participants are required to make the equation work by moving only one matchstick. Problem A is an example of the simplest problem type of this domain (Type A or standard type following Knoblich et al., 1999) and considered not to require restructuring (Öllinger, Jones, & Knoblich, 2008). Moving the vertical matchstick from the first value, "VI," to stand at the beginning of the same value (to make it "IV") yields the correct solution. In contrast, the constraint relaxation type Problem B is considered to require restructuring, because one needs to change the initial assumption that only the matchsticks from values can be manipulated (constraint relaxation type, CR3). In this case, the operator "+" can be decomposed and its vertical matchstick moved to make another "=" sign (VI = VI = VI). The "VI" in Problem A and the "+" sign in Problem B are the crucial elements that need to be changed for solution.

The Knoblich et al. study provides strong and objective evidence for the claim that a restructuring of the problem representation took place in Problem B. However, it did not answer the question of whether this change was sudden or gradual. There was a clear spike in paying attention to the important but previously ignored features in the final third of the allotted time, which would indicate a sudden restructuring. But this final period may have lasted minutes, as the solvers spent around five minutes on the problem. Thus, the restructuring might equally have been gradual, a continuous process over time.

Indeed, a recent eye tracking study by Ellis et al. (2011; see also Ellis & Reingold, 2014) found that participants started disregarding the irrelevant letter several seconds before they came up with the solution for anagrams, with viewing times on that letter decreasing gradually. Most intriguingly, both participant groups, those who experienced pop-out insight-like solutions and those who did not, displayed the same gradual accumulation of solution knowledge. There is the same recurring methodological caveat: in none of these studies (Ellis et al., 2011; Knoblich et al., 2001; Tseng et al., 2014) were the eye movement patterns analyzed on an individual level, but only aggregated across participants, thereby possibly obscuring sudden (or incremental) patterns present in individuals. If we want to make a statement about the temporal dynamics of restructuring, all trace data should also be analyzed on an individual level in order to be categorized as sudden or incremental.

The present study was designed to clarify what we call the cognitive hypothesis: do changes in the problem representation happen suddenly or incrementally? We aimed to provide a more fine-grained temporal analysis of the solution process by using ten time periods of equal length for our eye movement analysis instead of the three time periods used in the original Knoblich et al. study. If restructuring is a gradual process, we would expect that people who eventually solve correctly will start paying attention to the important aspect of the problem long before they find the solution (as found by Ellis et al., 2011). In contrast, if restructuring should be best characterized as a sudden process, we would expect the solvers to only pay attention to the important aspect immediately before they find the solution (as found by Knoblich et al., 2001). We also analyzed the temporal patterns at group level as well as at individual level.

Affective component of insight: the Aha! experience

With regard to the affective component of insight (Aha! experience), the question remains whether solvers are aware of the incremental or sudden nature of their solving process and whether they can report on it. The studies that assessed both the cognitive and affective components produced contradictory findings (Cushen & Wiley, 2012; Danek et al., 2018; Ellis et al., 2011). In Experiment 1b from Ellis et al., both participant groups, those who experienced pop-out insight-like solutions and those who did not, displayed the same gradual accumulation of solution knowledge. In the Cushen study, Aha! ratings were basically identical between solvers whose patterns had been categorized as insight-like and those whose patterns had been categorized as incremental. In contrast, the Danek study found a clear relationship between objective solution patterns and subjective solution experience, with sudden solution patterns leading to higher Aha! ratings than incremental patterns.

The second major goal of the study was to investigate what we call the affective hypothesis: is there a connection between solvers' subjective experience of the solution process and the objective temporal dynamics of restructuring (sudden or incremental)? To address this question, we measured the subjective experience after each trial using an Aha! phenomenology questionnaire adopted from previous studies (Danek, Fraps, von Müller, Grothe, & Öllinger, 2014; Danek & Wiley, 2017). If there is a connection between the affective and the cognitive component of insight, we would predict that the temporal course of the solution process should be reflected in solvers' subjective ratings of their solution experience. If a solution is found suddenly, solvers should also report a sudden emergence of the solution. Conversely, if the attention of eventual solvers gradually shifts towards the important feature, this should impact their phenomenological experience: they should experience less suddenness and less surprise than participants whose shift was more sudden, whereas the pleasure of finding a solution and the certainty about the correctness of their solution should not be impacted.

Matchstick arithmetic domain

In this experiment, we contrasted two matchstick problems with each other: one where prior knowledge is helpful, and one where it is not (see Figure 1). The matchstick arithmetic domain is ideally suited for eye tracking studies because each problem consists of the individual matchsticks, which allows for precise differentiation of fixations. There is also a full taxonomy of all four types of matchstick problems and their relative difficulty, which was theoretically derived from the representational change theory (Ohlsson, 1992) and empirically confirmed (Knoblich et al., 1999; Öllinger, Jones, & Knoblich, 2006, 2008). Here we chose the standard type (ST) problem (adopting the naming convention from Öllinger et al., 2008), which can be solved by manipulating a value (see Problem A in Figure 1). The problem is not trivial, but the solver will find the correct solution by trying out various ways of moving the matchsticks from one value to the other. Manipulating values conforms to prior knowledge of how equations can be solved and therefore this problem does not require any restructuring (Öllinger et al., 2008). In contrast, the other chosen problem (Problem B in Figure 1), the constraint relaxation type (CR3), requires a restructuring process because there are two constraints imposed by prior knowledge that need to be relaxed: the assumption that only values can be manipulated (instead, an operator needs to be changed) and the assumption that there is only one equal sign in an equation (instead, there can be two, creating a tautology).

Based on previous studies (Knoblich et al., 2001; Öllinger et al., 2008; Tseng et al., 2014), we expected that Problem A (ST) will be more frequently solved than Problem B (CR3), which involves restructuring. With regard to eye movements, we hypothesized that this would be reflected in a different pattern of attention allocation (see also, Knoblich et al. 2001; Tseng et al., 2014). In Problem A, participants can find the solution without fundamentally changing their representation. We expected them to focus on the values right from the beginning (an expression of the influence of prior knowledge, since, in equations, only the values can typically be manipulated), continuing throughout the entire solving period. Solvers will differ from nonsolvers insofar as they will focus more on the crucial value. For Problem B, it was also expected that all participants would initially focus on the values. Ultimately, solvers will shift their attention towards the crucial element (the operator) while nonsolvers will remain fixated on the values.

With regard to the affective component of insight, we hypothesized that the solution in Problem A would elicit less of the typical Aha! phenomenology, as measured by suddenness, pleasure and surprise (Danek et al., 2014), than the solution in Problem B (Knoblich et al., 2001). Further, we expected that the straightforward solution in Problem A would elicit more confidence about the solution being correct than the more difficult tautology solution in Problem B. This would provide evidence that the restructuring process necessary for successfully solving Problem B is closely connected to the Aha! experience.

We also expected a close correspondence between the cognitive and affective components of insight. The solvers who spend more time paying attention to the crucial element in Problem B, that is, who display incremental patterns of the restructuring processes, will also report less Aha! experience than the solvers who found the solution immediately after a sudden shift of attention towards the crucial element. We did not expect this pattern of results in Problem A because it does not feature the restructuring process.

Problem B is difficult and a good number of participants usually cannot find the solution in a reasonable period of time (Knoblich et al. 2001; Tseng et al., 2014). We decided to provide hints after a certain length of time to shorten the process of finding the solution, which may otherwise be too long for laboratory testing. Hints also present an additional check on the main assumption behind the reconstruction process. First, we expected a significant rise in paying attention to the elements mentioned in the hints. Second, should the piece of information provided by the hints be the crucial one, we should see a rise in solution rates after hints.

Method

Participants

Participants were 78 psychology students at Klagenfurt University who participated for course credits. They solved the first matchstick problem (Problem A) but we had to exclude five participants who were either familiar with this problem ($n = 1$) or had 33% of their eye data missing ($n = 4$), which left us with 73 participants on the first matchstick problem (7 male, M age $= 22.8 \pm SD$ age $= 6.2$ years). The second matchstick problem was solved by the same participants, but we had to exclude 12 participants because they were either familiar with the solution of that specific problem ($n = 2$), or their eye recordings were not reliable ($n = 10$), leaving us with 61 participants who tried to solve Problem B (5 male, M age $= 22.8 \pm SD$ age $=$ 6.5 years). All participants signed a written consent and the local ethics committee in Klagenfurt approved the study.

Our sample was considerably larger than those used in similar studies using eye movement recordings: $N = 24$ in Knoblich et al. (2001), $N = 38$ in Tseng et al. (2014), and $N = 32$ in Ellis et al. (2011). The effect size for the interaction between problem solving period and solving (solvers vs. nonsolvers) in the Knoblich study (p. 1007) is $\eta_p^2 = .42$ (calculated using F and df, see Lakens, 2013). In the Ellis study all participants solved the task, but one can estimate the effect size of the interaction between problem solving period and elements (relevant vs. irrelevant) – $\eta_p^2 = .69$ (p. 773). In both cases, the effects are so large than one should detect the effect even with just a few participants (fewer than 4 in each group with a 0.99 power). The Tseng study reports a large effect size for the difference between solvers and nonsolvers during the last third of the problem solving period ($d = 1.04$, which would mean that in the present study, a total of 60 participants is needed to achieve a power of 0.99 given the common significance level of 0.05). There are no studies that could be used to estimate the effect size for the interaction between eye movement patterns and Aha! experience (e.g., the Ellis study does not provide the necessary information in addition to using a different measure of insight phenomenology).

Stimuli, task and design

There were two problems. In the first (Problem A (ST) in Figure 1), participants needed to move a matchstick from a value to solve the problem (moving the I in VI to stand at the front makes IV, which together with the other IV gives the result, VIII). The second problem (Problem B (CR3) in Figure 1) involved changing the operator (+ becomes =) for the correct solution, thus creating a tautology (VI = VI = VI). Therefore, the "VI" in Problem A and the "+" sign in Problem B are the crucial elements that need to be changed for solution.

We first asked whether participants knew the solution to the first problem (Problem A). If so, this problem was skipped. They were then instructed as follows: "Here is a mathematical equation presented in Roman numerals. It is incorrect. Your task is to change the equation in such a manner that it is correct. You are allowed to move only one single matchstick. This picture helps you to return to the initial equation at any time. You have five minutes to think about the solution. Please let us know as soon as possible when you think you have the solution." No hints were provided for Problem A. The second problem (B) was presented with the same instruction. If participants did not solve Problem B within five minutes, we provided hints. The first hint, after five minutes, was: "You can change the operators, too." If participants did not solve the problem two minutes after the hint (7 minutes in total), they would receive a final hint: "You can also change only the operators". The last hint removes the possibility of using a part of an operator and adding it to a value. After further two minutes (nine minutes in total), the experimenter stopped the experiment.

The problems were always given one after the other (first A, then B), following the procedure in Knoblich et al. (2001). The solution was shown only after both problems were completed. The participants were instructed to press space bar as soon as they thought that they had found a solution to the problem. They would then say the solution out loud. If the solution was not correct, they continued to work on the same problem. If the solution was correct, the participants (only solvers) were given an Aha! phenomenology questionnaire (adopted from Danek et al., 2014; Danek & Wiley, 2017), addressing the following four dimensions of the Aha! experience: suddenness of the solution (*"The solution came to me… stepwise/suddenly"*), surprise (*The solution came to me… surprisingly/expectedly"*), certainty (*"I felt about the solution… uncertain/certain"*), and pleasure (*"At the moment of solution, my feelings were… pleasant/neutral"*). Answers were given on a 5-point Likert scale (no numbers shown) with only the extremes being labeled (see text in italics for the labels). After answering the Aha! questions, the participants were presented the next problem.

Participants were tested at the laboratory of the Department of Psychology in Klagenfurt. They were given instructions and a couple of warm-up problems (multiplication and division of two numbers[1]).

Eye movements

The whole experiment was presented on a 19-inch TFT monitor with a resolution of 1280 × 1024 pixels. Eye movements were recorded with the SMI RED250 mobile eye tracker with a sampling rate of 250 Hz and an accuracy of 0.4 degrees. All individual elements in the equation were taken as Areas of Interest (AOIs). In Problem A, all AOIs had the same dimensions of 152 × 157 pixels except the "result" which had the dimensions of 211 × 157 pixels (the result included an additional element compared to other AOIs which resulted in a bigger AOI). In Problem B, all AOIs had the same dimensions of 152 × 157 pixels. We set the minimum fixation duration at 5 ms with a fixation radius of 75 pixels (2 degrees). We excluded fixations that lasted for less than 100 ms (see also Knoblich et al., 2001).

Eye movements analysis

Due to the fact that every participant needed a different amount of time to complete the problems, any of the analyses to follow are in percentages of total individual solution time. For a comparison between participants who

---

[1] The simple arithmetic tasks are useful for checking the eye tracking equipment, as well as providing participants with confidence as the problems are easily solvable. In this particular context, they may have led participant to further fixate on values, which would in turn suppress solution rates in Problem B (which needed focus on operators for solution). We do not believe that this is the case as the solution rates of Problem B are almost identical to those in Knoblich et al 2001 study.

took different lengths of time to solve a task, the entire problem solving period was divided into ten equal intervals with each interval representing 10% of the total solution time of each individual solver (see Bilalić et al., 2008; and Knoblich et al., 2001, for a similar procedure of analyzing eye movement data). Unlike Knoblich et al. (2001), we did not differentiate between long and short fixations. Long fixations inevitably comprise most of the fixation time, which leaves little material for the analysis of short fixations. The analyses restricted to long fixations produced essentially the same results as presented here (the analysis of long and short fixations can be obtained upon request).

The analysis of relative problem solving time used in the main text enables to compare solvers who needed vastly different amount of time (e.g., solved very quickly or needed almost five minutes). The downside of this relative time approach is that the same amount of time (e.g., 30 s) needed for finding the solution after the reconstruction, when the participant focus on the important elements, will end up in different bin if the time needed to find the solution was vastly different (e.g., it could end up in the 5th to 10th bin when the total time is 60 s, or 9th and 10th bin if the time is 290 s). We provide an additional analysis to demonstrate that this is not the case. We run the same analysis on the relative bins as in the main text but use the time needed to find solution for the solvers as a covariate. If the total time needed for solution is an important factor for how much overall the solvers spend on the individual elements of the problems, the covariate solution time will be significant. If the solvers differing in time needed for solution were contributing differently to the individual bins, the interaction between bins and time on the individual element should be significant. This ANCOVA is conducted on solvers only as they varied in the solution time whereas the nonsolvers always needed the fully allocated time (5 minutes).

We also report the results by using the absolute instead of the relative solution time (see Appendix A, Figures A3 and A4). We use 5-second bins and move backwards from the solution to the beginning of the problem solving. This absolute time approach suffers from the same problem as the relative time approach – bins may reflect vastly differing stages of the problem solving process due to the differing time needed to solve the problem. More importantly, the missing values in certain bins make it difficult to connect the eye movement data with the affective ratings in further analysis (e.g., ANOVA or ANCOVA).

Results and discussion

Accuracy (success rate)

Figure 2 shows the success rate on both matchstick problems over time. Participants were better at solving matchstick Problem A (70% after

5 minutes) than matchstick Problem B (43%). This was formally confirmed by logistic regression on the success rate in the first five minutes – solving Problem A was more likely than solving Problem B ($b = 1.14$, SE = .36, $z = 3.5$, odds-ratio = 3.1, p = .002; Nagelkerke R^2 = .10).

Participants were better at solving the standard type Problem A (ST) than the constraint relaxation type Problem B (CR3) during the first five minutes (before the hints).

The hints played a role in the solving process of Problem B (see Appendix B, Figure B1). After receiving the first hints, an additional 15% of participants solved the problem in the next two minutes, and the last hint added another 23% solvers. Both hints helped additional 38% participants to solve Problem B. In the end, 80% participants solved Problem B after two hints and 9 minutes.

Standard type Problem A – eye movements

Figure 3 shows how much time the solvers and nonsolvers spent on every single element within the problem. At the beginning, most of the time (20%–30%) was spent on the "result" and the crucial element (VI) by both solvers and nonsolvers. As time passed, the solvers spent a gradually increasing amount of time looking at the crucial element of the equation. Shortly before solution, at the 90% interval, the differences between how much time solvers and nonsolvers were spending on the crucial element became more pronounced. Solvers were spending more time on the crucial element while the nonsolvers remained on the same level as in previous time periods. In contrast, nonsolvers spent somewhat more time than solvers on the other elements (e.g., result and plus sign).

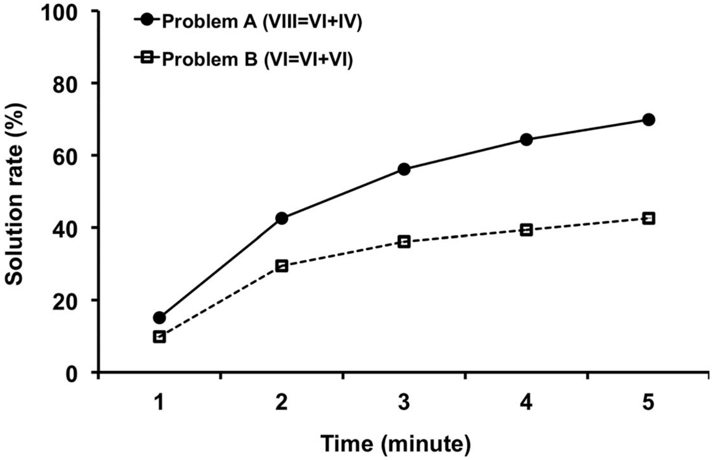


Figure 2. Cumulative solution rate in percentages for Problem A and Problem B.

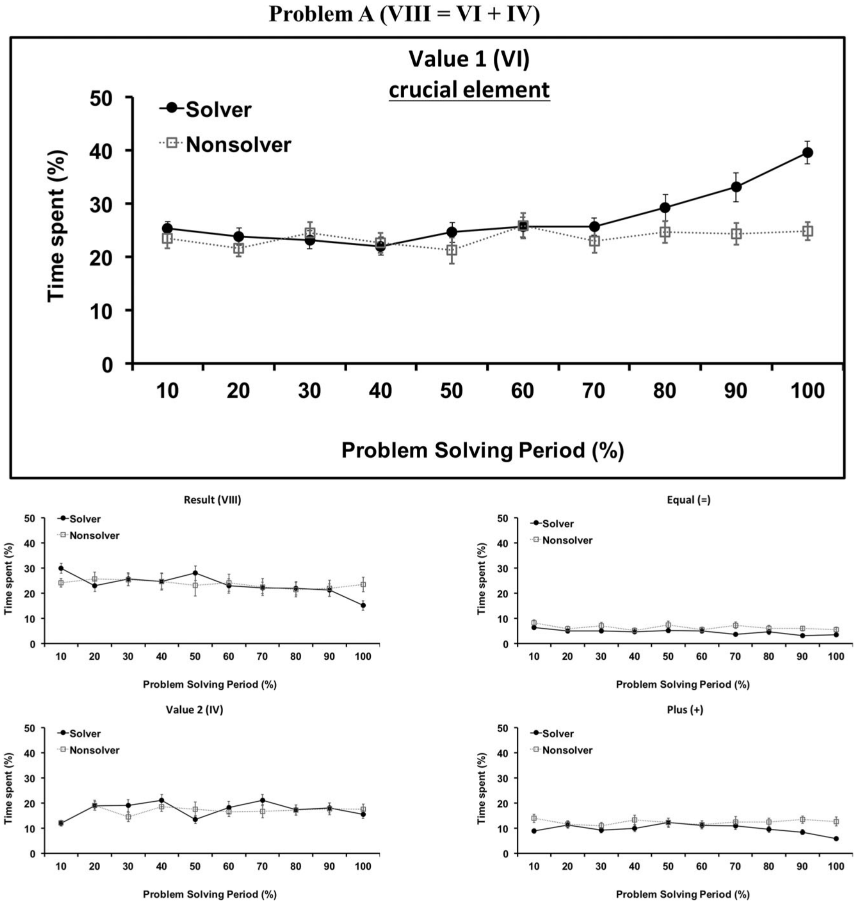


Figure 3. Percentage of time spent on elements of Problem A (ST) over the entire problem solving period (5 minutes).

To formally test these observations, we looked for the interaction between time period (10 bins) and group (solvers/nonsolvers) in a two-way ANOVA conducted for each AOI separately. Before conducting five two-way ANOVAs, as a way of avoiding the problem of testing multiple AOIs, we made sure that there was a significant difference between the AOIs in a three-way ANOVA which includes AOIs as an additional factor (grouped in two categories, values and operators). For the sake of brevity, we present these results in Appendix A. As expected, only the interaction between group (solver/nonsolver) and time period on the crucial element (value 1, VI, – see Figure 3) was significant (interaction group x time: $F(9,639) = 2.65$, $MSE = 365.6$, $p = .005$, $\eta_p^2 = .04$). This confirms that solvers and nonsolvers had significantly different eye movement patterns across the problem solving period on this particular element. That was not the case for any other equation element as the interactions between time period and group were not significant.

These results were not a product of differing time the solvers spent on the individual elements of the problems. When the time needed for solution was added as a covariate, its main effect and its interaction with bins was not significant for either the crucial element or other problem elements. The results of the absolute time analysis are presented in Figure A3 in Appendix A.

Participants mostly paid attention to the values ("IV" and "VI") and result ("VIII") at the beginning, and not to the operators ("+" and "="). Eventually, when they directed more attention towards the crucial value ("VI" – top panel), they found the solution. Error bars present standard error of mean (SEM).

Constraint relaxation type Problem B – eye movements

Figure 4 shows how much time solvers and nonsolvers spent on each of the five different elements of Problem B over the problem solving period without hints (the first five minutes). As with Problem A, we find that participants initially mostly paid attention to the values – over three times more than to the operators. Unlike in Problem A, a focus on the values is an indication that the problem induced a blockage, a fixation and possibly an impasse, as the solution cannot be found within the values. Participants who failed to find the solution maintained this mental representation, as evidenced by their paying consistently more attention to the values than to the operators over the course of the problem solving period. The solvers, in contrast, were able to overcome this pattern after some time, managing to break through the initial inappropriate mental representation and gradually switching their attention towards the crucial operators.

If we look at the two individual operators[2] in Figure 4, we can see that the main difference between solvers and nonsolvers is on the crucial element ("+"), but not on the equal sign. Solvers spent increasingly more time on the crucial element as time passed. This pattern was already evident in the middle of the problem solving period, around 60%, and became more pronounced with additional time. It seems as though the "result" was the one other element that showed an opposite pattern – solvers were spending less time on it as the problem solving period went on. In contrast,

---

[2] As for the values, the first value attracted a lot of attention, more than the second (see Figure 4) although they were the same (both VI). A possible reason for this could be that the same value (VI) was used in the previous problem and that it was the crucial element there, which needed to be changed for successful solution. It is therefore possible that the initial focus on the first value (twice as much as on the second) is a carry-over from the previous problem.

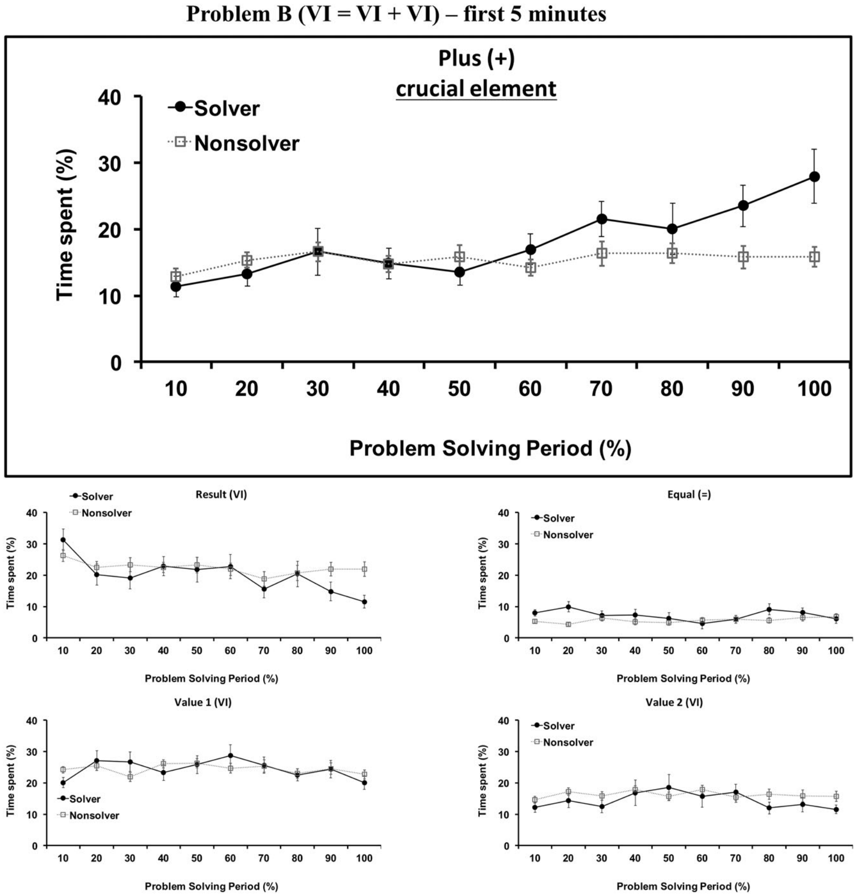


Figure 4. Percentage of time spent on elements of Problem B (CR3) over the first five minutes of the problem solving period (i.e., without hints).

nonsolvers divided their attention more equally across all problem elements over the course of the problem solving period.

We confirmed this pattern of results when we conducted a two-way ANOVA for each individual equation element separately (as with Problem A, we first conducted a three-way ANOVA including AOI as a factor and established that there is a difference between AOIs – see Appendix A). Solvers spent more time than nonsolvers on the plus sign in the second half of the problem solving period, which resulted in the significant interaction between group and time ($F(9,531) = 2.73$, MSE $= 334.8$, $p = .004$, $\eta_p^2 = .04$). The same interaction between time and group was not significant for the "result" ($F(9,531) = 1.5$, MSE $= 282.8$, $p = .15$, $\eta_p^2 = .02$). Other two-way ANOVAs with group and time as the main factors on the other individual equation elements did not produce significant interactions.

Again, the time needed to solve the problem among the solvers did not influence the pattern of results. When the time needed for solution was added as a covariate (ANCOVA on bins among the solvers), its main effect and its interaction with bins was not significant in the crucial element (covariate time needed for solution $F(1,24) = .7$, $MSE = 247$, $p = .41$, $\eta_p^2 = .03$; Interaction time needed for solution and bins $F(9,225) = .7$, $MSE = 149.9$, $p = .69$, $\eta_p^2 = .03$). We also found no influence of the time needed for solution on the other problem elements. The results of the absolute time analysis also confirm the incremental pattern of solvers (see Figure A4 in Appendix A).

Given the predictably low solution rate in Problem B after the initial time period, we also provided two hints. As analyzed in Appendix B, both hints about the possibility of using operators, given at the 5 and 7 minute marks respectively, inevitably drew more attention to the plus and equal signs and away from the values. Eventual solvers spent more time on the operators, especially the equal sign, as time passed. At the same time, they disengaged from irrelevant elements such as the values (VI).

Participants focused on the values ("VI" and "VI") and the result ("VI") at the beginning, but the solvers slowly switched their attention towards the operators ("+" and "=“). The solvers had already started paying attention to the crucial element ("+" – top panel) at the 60% interval. Error bars present SEM.

Aha! phenomenology questionnaire

Whenever participants solved a problem, we asked them for subjective ratings of suddenness, surprise, pleasure and certainty. Figure 5 presents the ratings from solvers of Problem A and Problem B during the first five minutes (that is before the hints, so that both problems are comparable). As expected, participants felt that the solution in Problem B was more sudden than the solution in Problem A (paired *t*-test with subjects who solved both problems for suddenness: $t(19) = 4.19$, $p < .001$, Cohen's $d = .94$). Solving Problem B elicited a stronger feeling of surprise and pleasure but neither difference was statistically significant (surprise: $t(19) = 1.83$, $p = .08$, $d = .41$; pleasure:: $t(19) = 1.72$, $p = .10$, $d = .38$). Finally, participants were more certain that they had found the right solution in Problem A than in Problem B, but this difference was also not significant ($t(19) = 1.37$, $p = .19$, $d = .31$). The trends were not statistically significant most likely because the majority of participants did not solve Problem B within the first five minutes. Once the solvers after hints were taken into consideration, the subjective experience of surprise was also significant (see Appendix B).

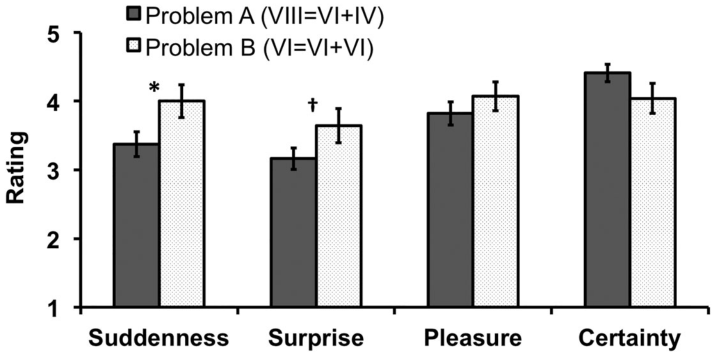


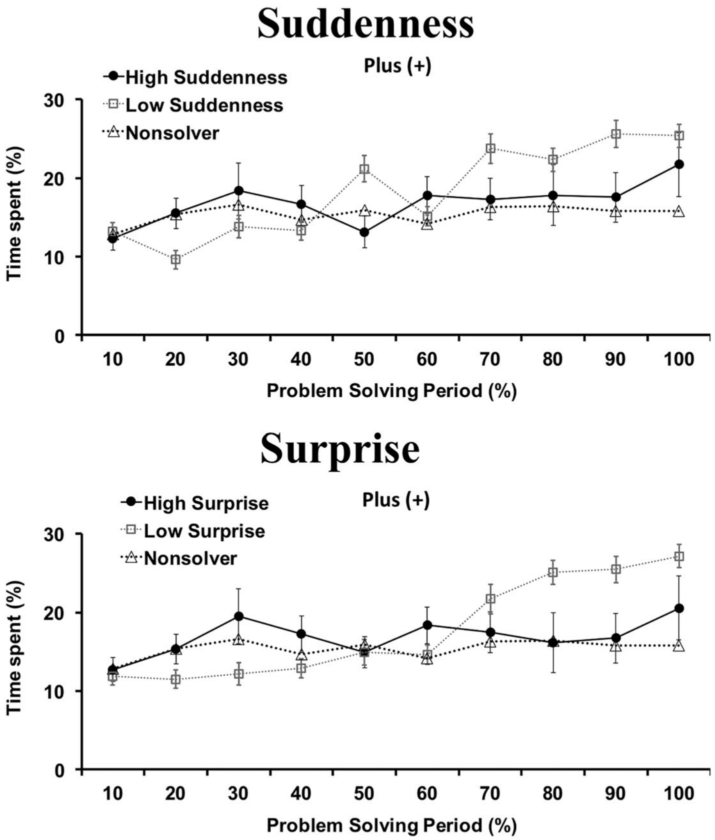
Figure 5. Aha! phenomenology questionnaire over the first five minutes. Average self-report ratings of all participants who solved both Problems A and B (first five minutes), with regard to how they experienced their solution process. * $p < .05$; † $p < .10$, paired *t*-test. Error bars present SEM.

Interaction between cognitive and affective component

To test our affective hypothesis, we checked whether the patterns of attention allocation were connected with the subjective experience of suddenness and surprise. We did this by including the rating as a covariate in the ANCOVA, which also included AOI and time (group was no longer applicable because only those participants who had solved the problem were asked about their solution experience with the Aha! phenomenology questionnaire). As with the previous analyses of eye movements, we first included all five AOIs in the analysis before we proceeded to check for the interaction on each separate AOI (see Appendix A).

In the standard type Problem A, we could not find any association between eye movement patterns and subjective ratings of suddenness, surprise, pleasure, or certainty. Even for individual AOIs, such as the crucial element ("VI" – see Figure 1), there was no interaction between the allocation of attention to the element over time and the subjective solution experience. Problem B, however, displayed a different pattern of results. Depending on when participants started to look at operators or values, their Aha! phenomenology ratings changed. The sooner solvers started paying attention to the crucial element ("+"), the less sudden ($F(9,342) = 1.93$, $MSE = 284.6$, $p = .046$, $\eta_p^2 = .05$) and less surprising ($F(9,342) = 2.54$, $MSE = 368.9$, $p = .008$, $\eta_p^2 = .06$) their solution felt.

In order to illustrate the interaction with suddenness and surprise in the ANCOVA analysis, we grouped the solvers into two groups, based on their ratings in the Aha! phenomenology questionnaire. The "High Suddenness/Surprise" group experienced the solution as very sudden/surprising (ratings 4 and 5 on a scale from 1 to 5), and the "Low Suddenness/Surprise" group


Suddenness

Surprise

Figure 6. Interaction between the cognitive and affective component in insight problem solving – group analysis, Problem B.

did not find the solution that sudden/surprising (ratings 1, 2, and 3). We then plotted their allocation of attention to the crucial element (see Figure 6). The subjective experience of the solvers in Problem B corresponded with how their solution process unfolded. As they started paying attention to the crucial element earlier, they also rated their solutions as emerging less suddenly and less surprisingly.

Participants with low subjective ratings of suddenness (top panel) and surprise (bottom panel) began to spend more time earlier on the crucial element (+) than participants with high subjective ratings. Nonsolvers are shown as baseline. Error bars present SEM.

Individual analysis

The previous analyses indicated that the restructuring process in Problem B (operationalized as the switch in attention away from the irrelevant values towards the crucial element) happens in a gradual manner, at least on an aggregate level. We also looked at the temporal dynamics of attention allocation in individual participants (see also Cushen & Wiley, 2012). Using independent raters who judged the patterns of changes in attention allocation towards the crucial element, we found that 54% (out of 26) of Problem B solvers within the first five minutes had "incremental patterns "with a gradual increase over time (for details of the whole procedure, see Appendix C). There were also a number of "sudden patterns" (27%), which displayed one sharp/ sudden increase across the problem solving period (the rest of the participants had patterns that were not possible to unambiguously classify). The example eye movement video for sudden and incremental solvers have been uploaded (they correspond to the patterns depicted in Figure C1, in Appendix C).

We then used the two categories of Problem B solvers (within the first five minutes) to check for the interaction between the cognitive and affective component. The solvers whose patterns of attention allocation were categorized as "sudden" gave higher suddenness ratings than solvers with incremental patterns (see Appendix C and Figure C2). This corroborates the ANCOVA results (see Figure 6) that the attention allocation patterns on the crucial elements are indicative of the subjective solving experience.

General discussion

This study addressed the question of whether the restructuring underlying insight problem solving happens suddenly or gradually (cognitive hypothesis) and whether solvers can actually perceive and report on these differences in temporal dynamics (affective hypothesis). At first glance, the present eye movement data supports the idea that restructuring happens gradually, because on an aggregate level, solvers paid increasing attention to the crucial element long before they actually found the solution. On an individual level, however, both sudden (27% of all solvers) and incremental (54%) looking patterns were found. Most importantly, there was a connection between objective and subjective measures of insight because solvers who started paying attention to the crucial aspects of the problem earlier also experienced their solution as less sudden and less surprising.

The hint analysis (Appendix A) further supports how difficult it is to find the solution even with the correct information. The hints did indeed change the attentional patterns as the participants started focusing more on the operators (see Figure A2 in Appendix A). However, this did not necessarily help them to immediately find the solution. The solvers consequently spent considerable time trying out different solutions that involved the operators. The nonsolvers even eventually returned to their previous unfruitful focus on the values.

Differences between the two matchstick arithmetic problem types

The standard type Problem A was solved more often than the constraint relaxation type Problem B. This difference in solution rates supports the theoretical assumption of the representational change theory (Knoblich et al., 1999; Ohlsson, 1984, 1992) that a restructuring process would be required for the solution of Problem B, but not for Problem A. Further evidence that Problem B triggers an initial incorrect problem representation (focus on the values) that needs to be restructured (focus on the operators) comes from the eye movement analysis, which revealed different patterns of attention allocation for the two problems. Right from the beginning, participants solving Problem A spent most of their time on the values, which for this problem included the crucial element needed for solution. Participants also focused on the values (which in this case did not include the crucial element) at the beginning of Problem B, which produced fixation. However, solvers then managed to change their problem representation, as evidenced by their gradually increasing the focus on the operators. The initial fixation on the wrong elements and the necessity of restructuring made Problem B considerably more difficult than Problem A. Together, these eye movement patterns support the main assumptions of the representational change theory of insight (Knoblich et al., 1999; Ohlsson, 1984, 1992), such as incorrect initial problem representation, fixation, and restructuring.

The subjective Aha! phenomenology ratings showed that Problem B also evoked a different solution experience from Problem A, confirming a link between the subjective feeling and objective measures of the insight solution process such as solution rates and eye movements. Solvers felt that the solution came more suddenly for Problem B than for Problem A. Similar trends were found for surprise.

Cognitive component of insight: restructuring

The use of a new analysis method, which looks into fine-grained temporal patterns of eye movements, allowed us to address the question about the temporal nature of the restructuring process apparent in Problem B. In the present study, the restructuring process, operationalized as a switch in attention towards the crucial element, started long before the solution was found, at least when the data was analyzed on an aggregate level. This indicates that the information needed for solution accumulates gradually, before solvers become aware of the solution, as also reported by Ellis et al. (2011). It should be noted that the sudden patterns were also revealed through the analysis of individual looking patterns. Although about half of all solvers incrementally allocated attention towards the crucial element, there were also about a quarter of solvers who underwent a sudden shift of attention to the crucial element. The estimates of the sudden patterns are different from other studies employing individual analysis (14% in Danek et al., 2018 and 68% in Cushen & Wiley, 2012). This is not necessarily surprising, given that both studies employed subjective ratings of problem elements instead of eye movements, as well as different domains (magic tricks and Triangle of Circles Problem, respectively).

The prevailing gradual manner of the restructuring process, on both the group and individual level, and on both pre- and post-hints analysis, speaks against one of the main assumptions of insight theories (Davidson, 1995; Duncker, 1926). Restructuring is not necessarily always a sudden process. This finding is, however, in accordance with the “insight stage” of Bower’s theoretical model of intuition (Bowers, Regehr, Balthazard, & Parker, 1990) where information is gradually accumulated until it becomes consciously available.

Affective component of insight: Aha! experience

With regard to the affective hypothesis, we found that participants seemed to be well aware of how their solving process unfolds. Solvers with incremental restructurings reported a diminished Aha! experience, which means that differences in the temporal dynamics of the restructuring process were reflected by differences in the Aha! phenomenology. As shown in Figure 6, the sooner solvers started paying attention to the crucial element, the less sudden and surprising the solution felt to them. We infer from this that there is a link between the objective temporal course of the solution process and its subjective perception.

This result is in accordance with a recent study that found a clear relationship between self-reported solution experience and actual solution patterns, with sudden solution patterns leading to higher Aha! ratings than incremental patterns (Danek, Williams & Wiley, 2018). Along the same lines, further supporting the idea of a connection between subjective solution experience and objective measures, another study (Salvi, Bricolo, Kounios, Bowden, & Beeman, 2016) demonstrated that solvers provide subjective ratings of their problem solving experience which can be quite accurate with regard to solution correctness. However, this finding stands in contrast to the studies which have reported a lack of sensitivity in self-reports of solution processes (Cushen & Wiley, 2012; Ellis et al., 2011).

One possible reason for this inconsistency is that the sensitivity of self-reports is dependent on the exact operationalization of the Aha! phenomenology in each study. There were several differences in the way that participants reported their solution experience in these previous studies. Typically, only one global Aha! rating is obtained with instructions that emphasize more than one phenomenological dimension. One problem arising from this procedure is that different solvers may rely on different dimensions when making their judgment. This important methodological issue is discussed in detail elsewhere (Bowden & Grunewald, 2018; Danek, 2018). In the Cushen and Wiley study (2012), participants rated whether the solution seemed surprising and sudden, in only one rating, without differentiating between these two dimensions. In the Ellis study (2011), they had to choose between four different options which described how the solution process felt and which differed mainly with regard to suddenness and awareness.

Clearly, these methods are different from the present study, where we decided to directly ask about how sudden, how surprising, how pleasant and how certain the solution felt, without pre-imposing any definition of the Aha! experience on the solver (see Danek, 2018, for some best-practice suggestions of how Aha! could be measured). This decision enabled us to disentangle the individual dimensions of the Aha! experience. Only the suddenness (and to some extent surprise) ratings, and not the pleasure or certainty ratings, differed as a function of when solvers started the restructuring process. A single global rating of Aha! would have obscured the fact that the standard type Problem A received lower suddenness (and to some extent surprise) ratings, but higher certainty ratings, than the constraint relaxation type Problem B. In an overall Aha! rating, the phenomenology of the two problems might not have differed at all. This supports the idea that the Aha! experience is a multi-dimensional construct (Danek et al., 2014) which should ideally be measured via its individual dimensions (Danek & Wiley, 2017).

Interaction between cognitive and affective components of insight

One possible account of the present findings involves the temporal difference between the moment when the restructuring happens and the moment when the solution is found. It is fair to assume that the restructuring takes place when the solver starts paying attention to the relevant elements of the problem, which were previously not considered. These elements then become a key part of the new, restructured problem representation. But after this restructuring, the solution may not follow immediately. Instead, several more thinking steps may be required to complete the solution, which in turn impacts the subjective solution experience. For example, instructing the participants to go outside of the nine-dot square in the nine-dot problem, which restructures their problem representation and should lead to the solution, does not immediately produce the solution (Weisberg & Alba, 1981).

Ohlsson postulated that the perceived suddenness of a solution is contingent upon the distance to solution, that is, how many thinking steps are still required once a potential solution element is identified (Ohlsson, 1984, 1992, 2011). Should the solvers happen to stumble upon the right solution immediately after restructuring, they would experience the typical Aha! feeling. But the more time solvers spend in the new, restructured mental representation looking for the solution, the more likely they are to gradually find the solution by examining one method of solving after another (see also Ohlsson, 1984). This may be what differentiates sudden from gradual solution patterns in the present study and explains why both patterns were found.

This idea is supported by a recent study that compared solutions to magic tricks which consist of only one crucial step with solutions that have several steps (Danek & Wiley, 2017). The multi-step solutions felt less sudden than single-step solutions, independent of the actual solution time or problem difficulty. Further support comes from another study (Danek, Wiley, & Öllinger, 2016), which found that problems for which at least three constraints had to be relaxed (Nine-Dot Problem) triggered less Aha! experiences than problems with only two constraints (Eight-Coin Problem and an operator type matchstick problem, CR2). It is plausible that the solution felt less sudden when more constraints had to be overcome, which in turn led to a diminished Aha! experience.

An explanation based on the temporal distance between the restructuring and the eventual solution would also explain the incongruent findings from Ellis et al.'s study (2011). People do not necessarily need to serially search for the solution after restructuring, but simply spending time on the solution may weaken the surprise effect that restructuring commonly has. Our solvers needed considerably more time to solve the matchstick problem than the participants in the study of Ellis et al., who found the solution to most anagrams within a few seconds. The short time between disregarding the distractor, which was necessary for restructuring in the anagram study, and arriving at the solution, may not have been enough for some participants to experience any Aha! moment.

Conceptual replication, limitations, and future directions

In the current climate of what some call "replication crisis" (Open Science Collaboration, 2015; but see also Stroebe & Strack, 2014 for a different view), it is important to highlight that our study conceptually replicates the seminal research by Knoblich et al. (2001). The values in the problems are different, but the problems have an identical structure (Problem B here, VI = VI + VI, and in Knoblich et al., III = III + III). It is therefore not surprising that both problems B yielded similar success rates. The differences are greater on Problem A (Problem A here, VIII = VI + IV, and in Knoblich et al., IV = III + III) as our participants are slower to start (15% after 1 minute compared to almost 90% in the Knoblich study) and some of them did not find the solution within the initial 5-minute period. A possible reason for this, besides the different participant groups, could be the micro-structure of the problem. In our study, the solution requires a change in the value (VI) of the equation. In the Knoblich study, not only does one need to change the value that constitutes the result (IV), but the repeated values of the equation (III + III) also leave fewer possibilities for manipulation, perhaps thus leading to higher solution rates. Similarly, we demonstrated essentially the same pattern of eye movements as in the Knoblich study, with the difference that we provided a more fine-grained analysis of the temporal dynamics.

A couple of possible limitations need to be considered. It may be likely that the solution in Problem B, which involved a tautology (VI = VI = VI), might have left participants wondering whether this was indeed the right solution. Besides the fact that problems with tautology solutions (i.e., CR3 type problems according to Öllinger et al.'s taxonomy, 2008) have been successfully used in several prior studies (Knoblich et al., 2001; Öllinger et al., 2008), the results on Problem B do not seem to be driven by participants being reluctant to announce the tautology. First, only a few participants asked whether the tautology was possible during the testing. Second, if they had indeed been reluctant to declare the unusual solution, the first hint about using the operators should have dispelled their doubts. There was no surge in the solution rate immediately after the hint, however. Only two participants solved Problem B within the first 20 seconds after receiving the hint. The same situation (two participants within first 20 seconds) was found after the second hint.

The use of matchstick arithmetic problems enabled us to build on a well-researched domain. We knew which problem type should elicit the restructuring process (Knoblich et al., 1999; Öllinger, Jones, & Knoblich, 2006, 2008) and could contrast it with a type which required no restructuring. We even had clear-cut predictions when it comes to the eye movement results (Knoblich et al., 2001). It is nevertheless only one domain and there is the question of generalizability to other insight domains. Future research should focus on insight problems where the reconstruction leads to a direct solution. The restructuring should draw the attention to the crucial element, but also leave no space for doubt about the solution. In the matchstick problems, it is necessary to try out different solutions even after one realizes that operators need to be changed. While only further research can answer whether the restructuring is incremental in these instances, there are several examples of where effects found in one insight domain generalize to other insight domains (Danek & Salvi, 2019; Kizilirmak et al., 2016; Webb, Little, & Cropper, 2016). Similarly, in the field of expertise, the findings from a seeming highly idiosyncratic domain such as chess are regularly found to hold in domains as diverse as radiology and sports (Bilalić, 2017, 2018).

Conclusion

The present study confirms the long-held notion (Knoblich et al., 2001; Öllinger et al., 2008) that some problems, such as the constraint relaxation type (Problem B), are more likely to trigger fixation within an initially incorrect problem representation. In contrast to standard type problems (Problem A), they subsequently require a change of that representation for solution (Ohlsson, 1992) and lead to a different solution experience (Aha! phenomenology). This supports the taxonomy of differentiating between insight and analytical problems (e.g., Gilhooly & Murphy, 2005; Weisberg, 1995) but we want to emphasize that this does not imply that these types of problems are always solved insightfully or analytically. As the simultaneous measurement of cognitive (eye movements) and affective (Aha! phenomenology questionnaire) components of insight in the present study showed, even nominal insight problems might not necessarily induce the typical Aha! experience in all solvers. Our study demonstrates the importance of the phenomenology-based approach in insight research, as postulated by Bowden et al. (Bowden & Grunewald, 2018; Bowden, Jung-Beeman, Fleck, & Kounios, 2005). The insight experience results from an interaction between the problem and the individual solver. Insight problems themselves are not guaranteed to elicit Aha! experiences (as already demonstrated by Danek et al., 2016; Webb, Little, & Cropper, 2016), but may provide more opportunities for this. The measurement of both cognitive and affective components is necessary if we want to capture the essence of the insight phenomenon. Similarly, the present study highlights the importance of analyzing solution patterns on the individual instead of the aggregate level (see also Cushen & Wiley, 2012; Fedor, Szathmáry, & Öllinger, 2015).

In summary, the present study replicates prior findings (Knoblich et al., 2001) that the solution process proceeds as postulated in the representational change theory (Knoblich et al., 1999; Ohlsson, 1984, 1992), but provides new evidence that the restructuring process does not always occur suddenly. Crucially, problem solvers are aware of the dynamics of their solving process and can accurately report the suddenness with which the solution emerged. We conclude that although restructurings may often happen gradually, it is only the sudden restructurings that lead to the full phenomenology of a sudden Aha! moment.

Acknowledgement

The help and cooperation from participants is greatly appreciated, as is Matthew Bladen's contribution in preparing the text.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by OeNB Jubil€aumsfondsprojekt 16449, Fortune Grant € F1354213, and DFG project BI 1450/1-2 to MB and DFG grant DA 1683/1-1, national grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, see www.dfg.de).

References

Ash, I. K., Jee, B., & Wiley, J. (2012). Investigating insight as sudden learning. *The Journal of Problem Solving*, *4*(2), 150–176. doi:10.7771/1932-6246.1123

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review*, *13*(1), 66–73. doi:10.3758/BF03193814

Ash, I. K., & Wiley, J. (2008). Hindsight bias in insight and mathematical problem solving: Evidence of different reconstruction mechanisms for metacognitive versus situational judgments. *Memory & Cognition*, *36*(4), 822–837. doi:10.3758/MC.36.4.822

Bilalić, M. (2017). *Neuroscience of expertise*. Cambridge: Cambridge University Press.

Bilalić, M. (2018). The double take of expertise: Neural expansion is associated with outstanding performance. *Current Directions in Psychological Science*, *27*(6), 462–469. doi:10.1177/0963721418793133

Bilalić, M., & McLeod, P. (2014). Why good thoughts block better ones. *Scientific American*, *310*(3), 74–79. doi:10.1038/scientificamerican0314-74

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition*, *108*(3), 652–661. doi:10.1016/j.cognition.2008.05.005

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science*, *19*(2), 111–115. doi:10.1177/0963721410363571

Bowden, E. M. (1997). The effect of reportable and unreportable hints on anagram solution and the Aha! experience. *Consciousness and Cognition*, *6*(4), 545–573. doi:10.1006/ccog.1997.0325

Bowden, E. M., & Grunewald, K. (2018). Whose insight is it anyway? In F. Vallee-Tourangeau (Ed.), *Insight: On the origins of new ideas* (pp. 28–50). London: Routledge.

Bowden, E. M., Jung-Beeman, M., Fleck, J. I., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences*, *9*(7), 322–328. doi:10.1016/j.tics.2005.05.012

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*(1), 72–110. doi:10.1016/0010-0285(90)90004-N

Cushen, P. J., & Wiley, J. (2012). Cues to solution, restructuring patterns, and reports of insight in creative problem solving. *Consciousness and Cognition*, *21*(3), 1166–1175. doi:10.1016/j.concog.2012.03.013

Danek, A. H. (2018). Magic tricks, sudden restructuring and the Aha! experience: A new model of non-monotonic problem solving. In F. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas* (pp. 51–78). London: Routledge.

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). It's a kind of magic – What self-reports can reveal about the phenomenology of insight problem solving. *Frontiers in Psychology*, *5*, 1408. doi:10.3389/fpsyg.2014.01408

Danek, A. H., & Salvi, C. (2019). Moment of truth: Why Aha! experiences are correct. *The Journal of Creative Behavior*. doi:10.1002/jocb.380

Danek, A. H., & Wiley, J. (2017). What about false insights? Deconstructing the Aha! experience along its multiple dimensions for correct and incorrect solutions separately. *Frontiers in Psychology*, *7*, 2077. doi:10.3389/fpsyg.2016.02077

Danek, A. H., Wiley, J., & Öllinger, M. (2016). Solving classical insight problems without Aha! experience: 9 Dot, 8 coin, and matchstick arithmetic problems. *The Journal of Problem Solving*, *9*(1), 47–57. doi:10.7771/1932-6246.1183

Danek, A. H., Williams, J., & Wiley, J. (2018). Closing the gap: Connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychological Research*, 1–9. doi:10.1007/s00426-018-0977-8

Davidson, J. E. (1995). The suddenness of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 125–155). Cambridge, MA: MIT Press.

Duncker, K. (1926). A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). *Pedagogical Seminary and Journal of Genetic Psychology*, *33*(4), 642–708. doi:10.1080/08856559.1926.10533052

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*, 270. doi:10.1037/h0093599

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science*, *5*(2), 94–97. doi:10.1111/j.1467-9280.1994.tb00637.x

Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cognition*, *20*(3), 768–776. doi:10.1016/j.concog.2010.12.007

Ellis, J. J., & Reingold, E. M. (2014). The Einstellung effect in anagram problem solving: Evidence from eye movements. *Frontiers in Psychology*, *5*, 679. doi:10.3389/fpsyg.2014.00679

Fedor, A., Szathmáry, E., & Öllinger, M. (2015). Problem solving stages in the five square problem. *Frontiers in Psychology*, *6*, 1050. doi:10.3389/fpsyg.2015.01050

Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, *25*(4), 436–463. doi:10.1080/20445911.2013.779248

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, *11*(3), 279–302.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. doi:10.3758/BF03212979

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., … Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*(4), e97–510. doi:10.1371/journal.pbio.0020097

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. doi:10.1016/0010-0285(76)90015-3

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*(3), 374–419. doi:10.1016/0010-0285(90)90008-R

Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., & Richardson-Klavehn, A. (2016). Generation and the subjective feeling of “Aha!” are independently related to learning from insight. *Psychological Research*, *80*(6), 1059–1074. doi:10.1007/s00426-015-0697-2

Kizilirmak, J. M., Serger, V., Kehl, J., Öllinger, M., Folta-Schoofs, K., & Richardson-Klavehn, A. (2018). Feelings-of-warmth increase more abruptly for verbal riddles solved with in contrast to without Aha! experience. *Frontiers in Psychology*, *9*, 1404. doi:10.3389/fpsyg.2018.01404

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534–1555. doi:10.1037//0278-7393.25.6.1534

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, *29*(7), 1000–1009. doi:10.3758/BF03195762

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012

Kuhn, G., & Land, M. F. (2006). There’s more to magic than meets the eye. *Current Biology*, *16*(22), R950–R951. doi:10.1016/j.cub.2006.10.012

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition*, *17*(6–7), 925–944. doi:10.1080/13506280902826775

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. doi:10.3389/fpsyg.2013.00863

MacGregor, J. N., & Cunningham, J. B. (2009). The effects of number and level of restructuring in insight problem solving. *The Journal of Problem Solving*, *2*(2), 7. doi:10.7771/1932-6246.1062

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 623–634. doi:10.1037/0278-7393.12.4.623

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*, 238–246. doi:10.3758/BF03197722

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

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology Section A*, *56*(2), 351–382. doi:10.1080/02724980244000288

Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, *25*(2), 117–129. doi:10.1111/j.1467-9450.1984.tb01005.x

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester-Wheatsheaf.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York: Cambridge University Press.

Öllinger, M., Jones, G., & Knoblich, G. (2006). Heuristics and representational change in two-move matchstick arithmetic tasks. *Advances in Cognitive Psychology*, *2*(4), 239–253. doi:10.2478/v10053-008-0059-3

Öllinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, *55*(4), 269–282. doi:10.1027/1618-3169.55.4.269

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. doi:10.1126/science.aac4716

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248.

Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players: Parallel encoding of chess relations. *Psychonomic Bulletin & Review*, *8*(3), 504–510. doi:10.3758/BF03196185

Rothmaler, K., Nigbur, R., & Ivanova, G. (2017). New insights into insight: Neurophysiological correlates of the difference between the intrinsic "aha" and the extrinsic "oh yes" moment. *Neuropsychologia*, *95*, 204–214. doi:10.1016/j.neuropsychologia.2016.12.017

Salvi, C., Bricolo, E., Kounios, J., Bowden, E. M., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Thinking & Reasoning*, *22*(4), 1–18. doi:10.1080/13546783.2016.1141798

Smith, R. W., & Kounios, J. (1996). Sudden insight: All-or-none processing revealed by speed-accuracy decomposition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1443–1462. doi:10.1037//0278-7393.22.6.1443

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59–71. doi:10.1177/1745691613514450

Tseng, C.-C., Chen, C.-H., Chen, H.-C., Sung, Y.-T., & Chang, K.-E. (2014). Verification of dual factors theory with eye movements during a matchstick arithmetic insight problem. *Thinking Skills and Creativity*, *13*, 129–140. doi:10.1016/j.tsc.2014.04.004

Webb, M. E., Little, D. R., & Cropper, S. J. (2016). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in Psychology*, *7*, 1424. doi:10.3389/fpsyg.2016.01424

Webb, M. E., Little, D. R., & Cropper, S. J. (2018). Once more with feeling: Normative data for the aha experience in insight and noninsight problems. *Behavior Research Methods*, *50*(5), 2035–2056. doi:10.3758/s13428-017-0972-9

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*(2), 169–192. doi:10.1037//0096-3445.110.2.169

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 157–196). Cambridge, MA: MIT Press.

Wertheimer, M. (1925). Über Schlussprozesse im produktiven Denken. In M. Wertheimer (Ed.), *Drei Abhandlungen zur Gestalttheorie* (pp. 164–184). Erlangen: Verlag der Philosophischen Akademie.

Appendix A: Initial (value-operator) 3-way ANO(C)VA analyses and absolute time analysis

Cognitive component: eye movements

Standard type Problem A

As mentioned in the main text, before embarking on the analysis of the individual elements, we analyzed the problems by grouping the equation elements in either values or operators (Figure A1). We conducted a three-way ANOVA on the percentage of time spent with the factors group (solvers/nonsolvers), time (10 periods of problem solving), and AOI (value/operator) to check whether these observations were significant. Both solvers and nonsolvers paid more attention to the values than to the operators (main effect AOI: $F(1,639) = 976$, $MSE = 753{,}789.6$, $p < .001$, $\eta_p^2 = .93$), but the solvers seemed to pay more attention to operators and less to values than nonsolvers (significant interaction group × AOI: $F(1,639) = 7.47$, $MSE = 5774.4$, $p < .008$, $\eta_p^2 = .10$).

Constraint relaxation type Problem B

We followed the same procedure for Problem B (first five minutes only). Figure A2 presents the eye movement allocation on the values and operators over the whole course of the problem solving period. We conducted a three-way ANOVA with group (solvers, nonsolvers), time (10 bins) and AOI (values, operators) which yielded not only a significant main effect of the AOI ($F(1,531) = 591$, $MSE = 427{,}300$, $p < .001$, $\eta_p^2 = .91$), but also a significant three-way interaction ($F(9,531)$

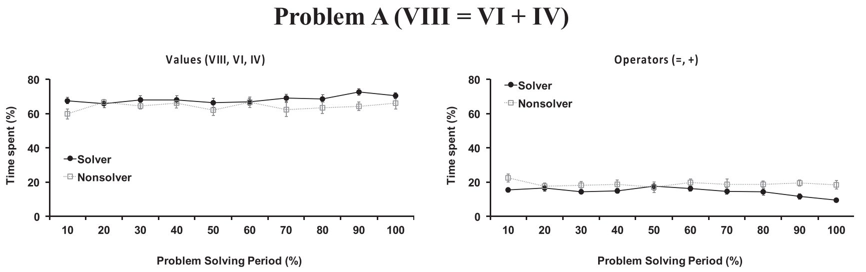


Figure A1. Percentage of time spent on the values (left) and operators (right) of Problem A (ST) over the five minutes of the problem solving period.

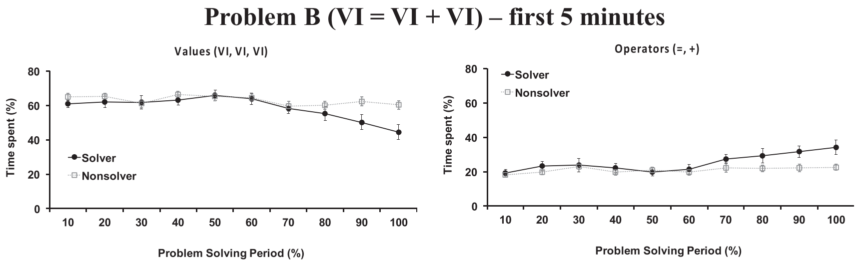


Figure A2. Percentage of time spent on the values (left) and operators (right) of Problem B (CR3) over the first five minutes of the problem solving period (i.e., without hints).

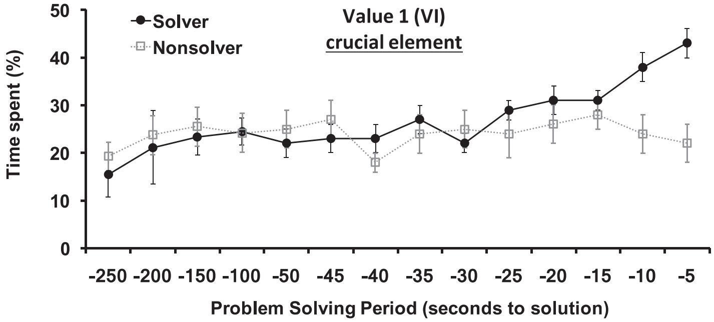


Figure A3. Percentage of time spent on the crucial element of Problem A (value 1 or "VI") in bins of five seconds, going backwards from the solution. The time before 50 seconds of the solution is given in 50 s bins.

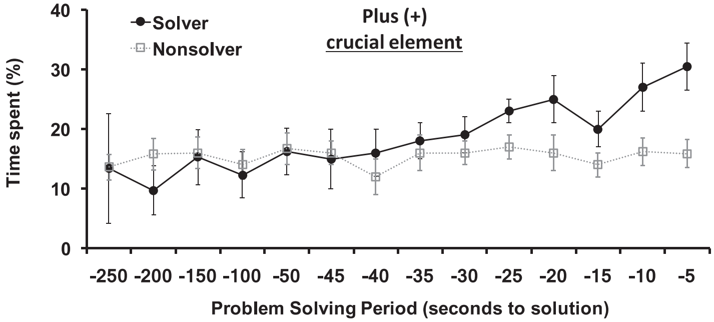


Figure A4. Percentage of time spent on the crucial element of Problem B (plus "+") in bins of five seconds, going backwards from the solution. The time before 50 seconds of the solution is given in 50 s bins.

$= 1.94$, MSE $= 647.7$, $p = .044$, $\eta_p^2 = .08$) as well as a significant two-way interaction between group and AOI ($F(1,531) = 7.67$, MSE $= 5534.9$, $p = .008$, $\eta_p^2 = .12$). In the main text we then demonstrated that the driving force behind the interaction was the crucial element (see Figure 4).

Interaction between cognitive and affective component

As with the previous analyses, before looking at the individual elements of a problem, we first conducted ANCOVA with time and AOI (operators vs value) as factors and rating as covariate. We were interested in the three-way interaction between time, AOI, and rating. None of the ratings produced any effects in Problem A. Problem B, however, displayed a different pattern of results. Depending on when participants started to look at operators or values, their Aha! phenomenology ratings changed. The sooner solvers started paying attention to the operators, the less sudden ($F(9,342) = 1.90$, MSE $= 664.5$, $p = .050$, $\eta_p^2= .05$) and less surprising ($F(9,342) = 2.32$, MSE $= 837.6$, $p = .012$, $\eta_p^2= .06$) their solution felt. No such patterns of results were found for pleasure or certainty. We show in the main text that the effect is driven by the differences on the crucial element ("+") – see Figure 6.

Absolute time analysis

We supplement the relative bin analysis presented in the main text with the analysis on the absolute time. The bins are now always constant time periods of five seconds moving backwards from the moment the participant found the solution (or first 300 seconds elapsed). Figure A3 shows the time spent on the crucial element in Problem A by solvers and nonsolvers. The differences are only visible in the last 10 seconds.

Figure A4 shows the time spent on the crucial element in Problem B by solvers and nonsolvers. The differences between solvers and nonsolvers are now visible already at around 25 s before the solution.

Appendix B: Hint analysis

Cognitive component: eye movements

There were two hints in Problem B, one after five minutes, and the other after nine minutes. Figure B1 shows that the hints helped the participants to find the solution.

The hints were supposed to draw attention towards relevant aspects of the problem. The first hint in Problem B was given after five minutes. Figure B2 shows the pattern of attention allocation among different equation elements across time for the participants who found the solution in the next two minutes (solvers) and those who did not (nonsolvers). The hint about the possibility of changing the operators worked as intended, and inevitably drew more attention to the plus and equal signs and less to the values. Please consult Figure 3 (panel with "+") in the main text.

This resulted in a similar amount of attention paid to the operators and values, unlike in the initial first five minutes (Figure 3). Eventual solvers started spending more and more time on the operators and less on the values, while nonsolvers divided their focus equally between values and operators over the problem solving period. This pattern of results led to a significant three-way interaction

Figure B1. Cumulative solution rate in percentages for Problem A and Problem B. Participants were better at solving the standard type Problem A (ST) than the constraint relaxation type Problem B (CR3). The hints after 5 and 7 minutes (indicated by the dotted lines in the graph) helped additional participants to solve Problem B. No hints were provided for Problem A.

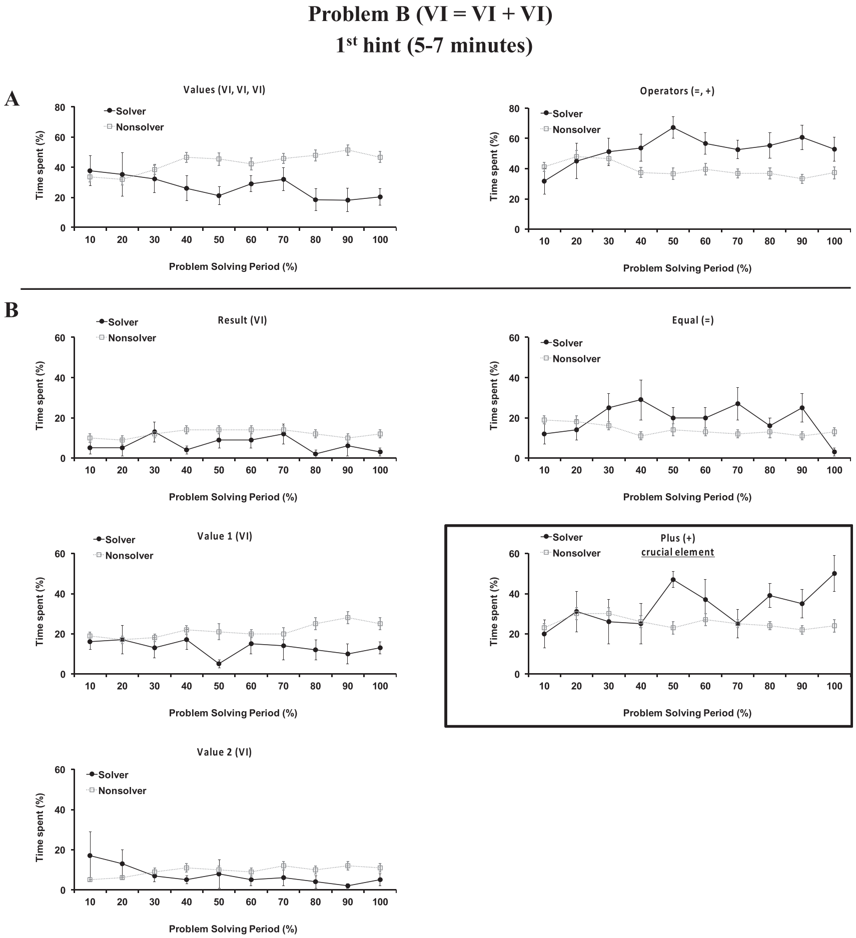


Figure B2. Percentage of time spent on elements of Problem B (CR3) after the first hint (5-7 minutes) for A) all values and operators together and B) individual elements. Participants focused equally on the values and the operators at the beginning, but then solvers slowly increased the amount attention that they directed towards the operators. Error bars present SEM.

group × time × AOI as measured by values and operands ($F(9{,}288) = 2.71$, MSE $=$ 1549.66, $p = .005$, $\eta_p^2 = .08$), as well as a significant two-way interaction between group and AOI ($F(1{,}288) = 16.78$, MSE $= 21{,}200$, $p < .001$, $\eta_p^2 = .34$).

When we looked into the equation elements separately, the interaction between time and group was unsurprisingly significant for the crucial element "+" ($F(9{,}288) = 2.99$, MSE $= 607.81$, $p = .02$, $\eta_p^2 = .08$). As time went by, solvers paid more and more attention to the plus sign, unlike nonsolvers. The consequence of this was that the solvers also disengaged from the values. The interaction between group and time was significant for the second value ($F(9{,}288) = 2.59$, MSE $=$ 236.78, $p = .007$, $\eta_G^2 = .08$), but not for the first value ($p = .71$). The interaction time x group effects were not significant for the "result" ($p = .86$), but were significant for the equal sign ($F(9{,}288) = 2.93$, MSE $= 448.73$, $p = .002$, $\eta_p^2 = .08$).

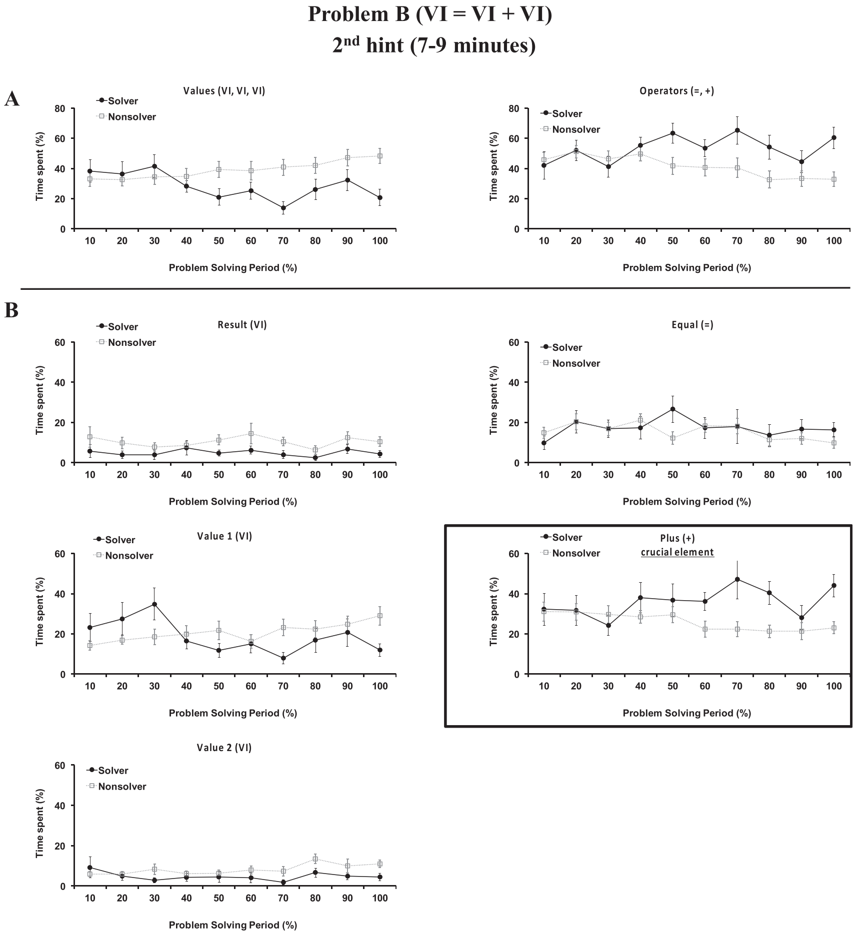


Figure B3. Percentage of time spent on elements of Problem B (CR3) after the second hint (7-9 minutes) for A) all values and operators together and B) individual elements. Participants focused equally on the values and the operators at the beginning, but then solvers slowly increased the amount of attention that they directed towards the operators. Error bars present SEM.

Similar effects were expected after the second hint, which came after the participants could not solve the problem within seven minutes. Figure B3 demonstrates that this was indeed the case. The second hint, about the possibility of changing only the operator signs, inevitably drew more attention to the operators and less towards the values, but this time without a significant interaction. As with the first hint, this led to a significant three-way interaction group x time x AOI as measured by values and operators ($F(9,216) = 3.21$, MSE $= 1916.17$, $p = .001$, $\eta_p^2 = .12$), as well as significant two-way interaction between group and AOI ($F(1,216) = 7.59$, MSE $= 15{,}994.23$, $p = .011$, $\eta_p^2 = .24$).

We then checked the individual elements for the interaction between solvers/nonsolvers across time. Solvers allocated increasingly more attention to the plus sign ($F(9,216) = 2.05$, MSE $= 606.43$, $p = .036$, $\eta_p^2 = .08$). On the other hand, solvers

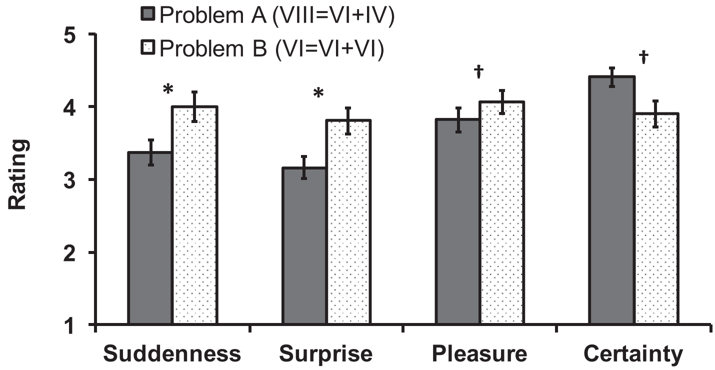


Figure B4. Aha! phenomenology questionnaire across the whole problem solving period. Average self-report ratings of all participants who solved both Problems A and B (all participants, including hints), with regard to how they experienced their solution process. * $p < .05$; † $p < .06$, paired *t*-test. Error bars present SEM.

disengaged from the first value over time whereas nonsolvers spent consistently similar amount of time inspecting the first value ($F(9,216) = 3.53$, MSE $= 768.23$, $p < .001$, $\eta^2_G = .13$). This time, the interaction between group and time on other elements, including the second value, was not significant.

Affective component: insight questionnaire

Figure B4 shows the subjective ratings of suddenness, surprise, pleasure and certainty when the participants who solved Problem B with the help of hints were included. As in Figure 4 in the main text, Problem B elicited more suddenness ($t(33) = 3.45$, $p < .001$, Cohen's $d = .60$), but the difference for surprise was also significant ($t(33) = 3.27$, $p = .002$, Cohen's $d = .56$). The differences between Problem A and Problem B were borderline significant for pleasure ($t(33) = 1.98$, $p = .057$, Cohen's $d = .34$) and certainty ($t(33) = 2.01$, $p = .053$, Cohen's $d = .34$).

Appendix C: Individual solution pattern analyses for Problem B solvers (within first five minutes)

This analysis was conducted only for Problem B, because our previous analyses had shown that this was the only problem where a switch in attention towards the crucial element (i.e., restructuring) had taken place. The patterns of changes in attention allocation towards the crucial element were categorized as being a sudden increase across time, an incremental increase, or other (this included flat, decreasing, i.e., less attention allocated to crucial element, or zigzag patterns with no clear dynamic in either direction). Adopting the methodology of previous studies (Cushen & Wiley, 2012; Danek et al., 2018), a line graph was created from the pattern of each participant who solved Problem B without hints ($n = 26$). Solvers after hints were not included in this analysis because there is evidence that receiving hints may bias the solution process towards the more incremental approach (Bowden, 1997; Cushen & Wiley, 2012; Davidson, 1995).

Three raters engaged in a visual analysis of the 26 graphs and made a (sudden increase, incremental increase, other) judgment for each graph. The two-way random intraclass correlation coefficient (absolute agreement), ICC (2, 2) was .83 with

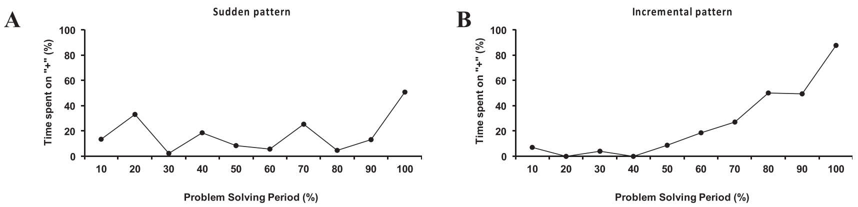


Figure C1. Actual examples of eye movement patterns on the crucial elemental in Problem B (+) that were categorized as A) sudden and B) incremental.

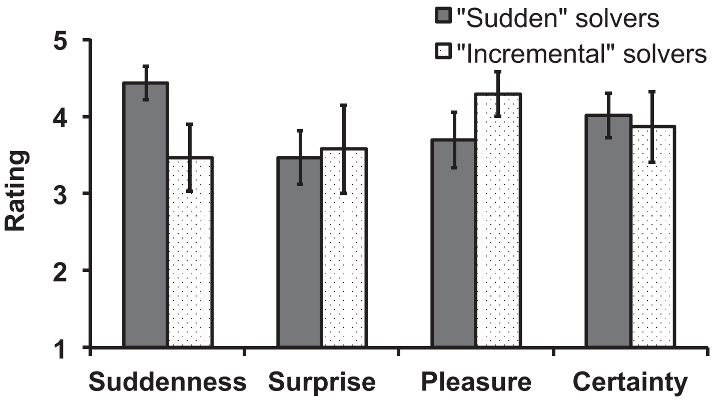


Figure C2. Problem B. Ratings of suddenness, surprise, pleasure, and certainty of solvers whose eye movement patterns were categorized as "sudden" or "incremental" (first five minutes). Error bars present SEM.

a 95% confidence interval of [0.67;0.92], indicating a good level of agreement (Koo & Li, 2016). Conflicting cases were resolved by a fourth rater. Figure C1 gives examples of patterns that were coded as sudden or incremental (the corresponding eye movement video have been uploaded).

There were 54% incremental solvers, 27% sudden solvers, while the rest (19%) could not be recognized unambiguously as one of the two patterns.

Finally, we supplemented the group analysis (see Figure 6) by checking for the ratings of suddenness and surprise of the individual solvers categorized as "sudden" or "incremental". This analysis of solvers' individual eye movement patterns showed that solvers whose patterns of attention allocation were categorized as "sudden" gave higher suddenness ratings than solvers with incremental patterns (Figure C2). A Welch's *t*-test for unequal variances showed that this difference did not quite reach the significance level ($t(16.25) = 2.02$, $p = .06$), probably due to the small number of participants who solved Problem B without hints ($n = 26$). The effect size, however, was medium to large (Cohen's $d = 0.74$, calculated using the pooled standard deviation of the two groups because SDs were unequal). Figure C2 also shows that there were no differences with regard to the level of surprise reported ($p = .86$). "Sudden solvers" also felt less pleasure than "incremental solvers" but this difference was not significant ($t(18) = 1.09$, $p = .29$, $d = 0.51$). There was no difference in the feeling of being certain about the solution between the two groups of solvers ($p = .66$).