Cognition xxx (xxxx) xxx-xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: http://ees.elsevier.com

# What causes the insight memory advantage?

Amory H. Danek *, Jennifer Wiley

Cognitive Division, Department of Psychology, University of Illinois at Chicago, 1007 W. Harrison Street MC 285, Chicago, IL 60607, USA

ARTICLE INFO

*Keywords*
Insight problem solving
Memory
Metacognition
Aha! experience
Restructuring
Confidence

ABSTRACT

Prior research indicates that solutions accompanied by an Aha! experience are remembered better than those missing this feeling of epiphany. The question for the present studies was whether this *insight memory advantage* for problem solutions is modulated by the affective component of insight (the strong feelings that typically accompany the Aha! experience), or by the cognitive component (the restructuring or representational change that occurs during insightful problem solving). In both studies, participants viewed a set of magic trick videos to generate solutions for how each trick was done, and memory for the generated solutions was tested after a week delay. They also indicated the extent to which they experienced an Aha! moment at solution along with other perceptions of their experience. In the second study, they additionally rated the relevance of five action verbs for each trick (including one that implied the correct solution) multiple times during solution as a measure of restructuring the problem representation. The explanation for the *insight memory advantage* that was best supported by the results is that it is the joint consequence of finding correct solutions, the subjective feeling that one has found a correct solution (certainty), and experiencing an emotional pleasurable reaction during the problem solving process that all contribute to better memory for the solution. However, it did not seem to rely on having reached the solution via a sudden restructuring process.

## 1. Introduction

Recent evidence indicates that insightful problem solving might leave a mark on memory: Self-generated solutions for which solvers report an Aha! experience are remembered better than those where this feeling of epiphany is missing (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013; Kizilirmak et al., 2016). However, the Aha! experience is a multidimensional construct that can emerge from a number of different factors (Danek et al., 2014a; Danek & Wiley, 2017; Webb et al., 2016, 2017). The intriguing question is why this *insight memory advantage* for problem solutions occurs and which factors are responsible for it. In particular, it is unknown whether the memory advantage is modulated by the affective component (the strong feelings that typically accompany the Aha! experience), or by the cognitive component (the restructuring or change within the problem representation that occurs during insightful problem solving) or whether both are required for the memory advantage to occur (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013; Gick & Lockhart, 1995; Kizilirmak et al., 2016). The overall aim of this research was to understand the contributions of affective and cognitive components of solutions to the *insight memory advantage*.

The idea that experiencing an insight could contribute to better memory for a solution is intuitively appealing and has been part of the insight narrative for a long time (see e.g., Dominowski & Dallob, 1995; Osgood, 1953; Scheerer, 1963; Woodworth & Schlosberg, 1954). Hebb asserted that “Whatever insight is, we now know that it continually affects the learning of the adult mammal” (Hebb, 1949, p. 163). In the very first experiments on insight, Köhler described that his chimpanzees were typically very quick in re-producing solutions that they had discovered earlier (Köhler, 1921). Dominowski and Buyer (2000) explored the same effect in humans using a set of puzzles generally assumed to require insight for solution. It was shown that correctly solved problems were nearly perfectly re-solved after one week in contrast to problems which had not been solved correctly. Further, the advantage was found to be specific to generating a correct solution, as participants who were simply told the correct solution did not show the same near-perfect memory for those solutions (Dominowski & Buyer, 2000). In a third experiment, the advantage disappeared when the connection between the problem and the solution was not meaningful but arbitrary (in this case, the given problems were cue words that had randomly been paired with solution words which had to be selected from a list, ostensibly using extrasensory perception). Based

* Corresponding author at: Department of Psychology, University of Heidelberg, Heidelberg, Germany.
*E-mail address:* danek@psychologie.uni-heidelberg.de (A.H. Danek)

https://doi.org/10.1016/j.cognition.2020.104411
Received 2 September 2019; Received in revised form 8 July 2020; Accepted 14 July 2020
Available online xxx
0010-0277/© 2020.

on these results, and consistent with Gestalt theory (Duncker, 1945; Köhler, 1921; Maier, 1945; Ohlsson, 1992; Wertheimer, 1959), they suggested that a reorganization or restructuring of information occurs during the generation of a correct solution, and this is what leads to better memory for solution information. Conversely, simply being told the correct solution does not require restructuring or changes in the solvers' problem representation, and therefore provides no advantage.

A subsequent study provided additional evidence that better memory for solutions may be linked to the restructuring that occurs in problem representations during insightful solution processes. Ash and Wiley (2008) explored the effects of solution processes on memory by using a measure of hindsight bias for initial problem representations, administered one week after the solution attempt. They demonstrated that hindsight bias in memory for the initial problem representation occurred for a set of puzzles that generally require representational change for solution (as they tend to mislead solvers into incorrect initial problem representations, Ash et al., 2012). Solvers remembered their initial representations as being more similar to the representations needed for correct solutions, demonstrating a hindsight effect. In contrast, no hindsight bias occurred in memory for initial representations on math problems which did not require any restructuring for solution. Further, to test for the differences in hindsight between correctly generating a solution and being shown a solution, all participants were shown the correct solutions to problems at the end of the problem set. No hindsight occurred for unsolved puzzles for which the correct solution had been shown. Only when people reached the correct solutions to the puzzles on their own was the memory effect seen. The interpretation was that the final problem representation that allowed for a correct solution overrode solvers' memory for incorrect initial problem representations. Consistent with the arguments made by Dominowski and Buyer (2000), when a solver is shown a correct solution it obviates the need to restructure the problem representation. A change in representation is no longer needed to know the correct solution. Thus, the lack of a hindsight effect from shown-solutions can be inferred to be related to a lack of restructuring. This study suggests a potential association between solution processes involving restructuring and better memory for solutions.

In contrast, other researchers emphasize the feelings associated with insight, or the emotional or hedonic component of the insightful solution process (Cosmelli & Preiss, 2014; Gick & Lockhart, 1995; Gruber, 1995; Topolinski & Reber, 2010), which was described by Poincare as "aesthetic emotion" (as cited in Topolinski & Reber, 2010) and by Thagard and Stewart as the "ecstasy of discovery" (Thagard & Stewart, 2011, p. 10). Still others point to the important role of a feeling of immediate confidence or certainty that is present even before the explicit verification of an idea, a feeling they refer to as an "intuitive sense of success" (Gick & Lockhart, 1995, p. 215). One study often cited in relation to these points is Auble et al. (1979) who suggested that it is the affective Aha! response that makes an insightful solution experience more memorable. In their study, they presented participants with initially incomprehensible sentences (e.g., "The haystack was important because the cloth ripped."). In the delayed-cue condition, each sentence was followed by a cue that revealed the meaning of the sentence only after a delay ("parachute"). In the immediate-cue condition, participants saw the same sentences with the cues embedded in them. The delayed-cue condition led to better recall for sentences, presumably because it caused an Aha! experience. Auble et al. suggested that Aha! experiences may facilitate recall either by the affective reaction promoting memory for the solution in semantic memory, or due to the distinctiveness of the Aha! reaction in episodic memory for the solution event or experience.

On the other hand, the Auble et al. results can also be explained in another way. One might argue that memory might have been improved via the cognitive restructuring component of insightful problem solving. Participants in the delayed-cue condition may have been more likely to develop an incorrect initial problem representation (e.g., "cloth" referring to a farmer's dress) that led to an impasse, or gap in knowledge that needed to be filled in order to reach solution, as compared to participants in the immediate-cue condition who may have been more likely to develop a correct initial problem representation ("cloth" referring to parachute). It may be that the sudden emergence of a new way of looking at the problem when given the delayed cue may have been responsible for the better memory for that solution. Although this is a plausible alternative explanation for the Auble et al. results, it could not be directly tested because this study collected neither measures of representational change nor of Aha! experiences during problem solving.

More recent studies have begun to establish a link between Aha! experiences and the *insight memory advantage* by collecting measures of participants' subjective solution experiences and utilizing magic tricks as new problem-solving stimuli (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013). To appreciate this new paradigm, consider the magic trick called "Match through Match". The magician holds one matchstick between the forefinger and thumb of each hand, hits them a few times against each other to show that they are solid and then quickly moves one matchstick through the other one. Although the act appears impossible without breaking one of the matches, at the end both matches are shown intact. An initial solution approach that many participants consider involves breaking one of the matches, yet that initial solution does not provide a satisfactory explanation for the unbroken matches at the end of the trick. The actual solution is that the magician imperceptibly raises one forefinger to create a gap which allows the other match to pass through. Finding this solution requires making a change in the representation from "breaking" a match to "raising" a finger. A new solution approach must be found and the initial assumptions must be overcome, just as in other insight problem solving tasks (Danek et al., 2014b). In addition, the solving of magic tricks has been shown to trigger strong Aha! experiences (Danek et al., 2014a; Danek & Wiley, 2017; Hedne et al., 2016).

In Danek, Fraps, von Müller, Grothe, & Öllinger, 2013, participants judged whether they experienced an Aha! moment after generating the solution for each trick, with no feedback on their solutions. When asked to recall their solutions two weeks later, an *insight memory advantage* was found: Solutions accompanied by an Aha! were recalled better than solutions for which no Aha! was reported. At the same time, Danek, Fraps, von Müller, Grothe, & Öllinger, 2013 found independent benefits from generating correct versus incorrect solutions and from Aha! experiences. Given that a large part of the artistry involved in magic is presenting the trick in such a way that it promotes an initial incorrect problem representation (Danek, 2018), it is possible that it was the change from an initial incorrect problem representation to a correct representation that led to the improved memory for the correctly generated solutions to the magic tricks.

Kizilirmak et al. (2016) reported similar memory results using a very different paradigm. Participants were given a perceptual problem solving task where they had to identify which object was pictured in a degraded image. Participants provided an Aha! judgment after generating their initial solutions. One week later, they were given a second chance to identify the objects for the same degraded images. Similar to the Danek, Fraps, von Müller, Grothe, & Öllinger, 2013 study, participants who had previously generated correct solutions accompanied by an Aha! experience were more likely to identify the object than participants who had previously generated correct solutions without an Aha! experience. Kizilirmak et al. (2016) also found that memory was enhanced for correctly generated solutions. Although this study did not examine recall when participants generated incorrect solutions (those were dropped from the analysis), it presented the correct solutions when participants failed to generate solutions within the time limit. Solvers were more likely to identify the objects on the second exposure on images for which they had correctly generated solutions over images for which they were shown the correct solutions. While attempting to perceive patterns in degraded images may not necessarily promote incorrect initial representations in the same way as magic tricks (although see Ludmer et al., 2011; Salvi et al., 2016; Schooler & Melcher, 1995), following the same logic as discussed by Dominowski and Buyer (2000), the memory advantage in the generation condition could also suggest that actively discovering a solution was more likely to affect internal representations. Generating a solution could alter the perception of features relevant for perceiving the object to a greater extent than simply receiving that solution information. By this reasoning, the memory advantage observed in the Kizilirmak study could also have plausibly been due to restructuring that occurred in the process of an initial problem solving attempt. Thus, the significant effect of correctly generating solutions on solution memory may reflect the contribution of the cognitive component of insightful problem solving to the *insight memory advantage*.

In contrast, Danek, Fraps, von Müller, Grothe, & Öllinger, 2013 attributed the independent contribution from the Aha! ratings (over and above the contribution of generating a correct solution) as being due to the affective component, and particularly the role of positive affect or pleasure as being responsible for the *insight memory advantage*. Even though the Aha! prompt provided in that study did not mention affect, joy, or pleasure at all, in an open-ended survey done at the end (reported in Danek et al., 2014a), solvers spontaneously described the Aha! moments that they had experienced as being marked by pleasure, enjoyment, or happiness. Participants may either intuitively perceive "pleasure" to be a hallmark of the Aha! experience, or it may be that the strong positive feelings caused by insightful solutions may "leak" into subjective ratings of the Aha! experience (as explained by the feelings as information theory, Schwarz & Clore, 1983, 2003). For either of these reasons, Aha! ratings may be reflective of the pleasurable positive affect that is thought to accompany the Aha! experience, as also proposed by Topolinski and Reber (2010).

Other evidence to support this interpretation comes from the study by Kizilirmak et al. (2016) which directly assessed the relation between Aha! and positive affect by including a pleasure rating (using smileys). They demonstrated that pleasure ratings were higher for solutions accompanied by an Aha! experience compared to solutions without one. Other studies have since corroborated this finding of a strong relation between pleasure and Aha! ratings for individual problem solutions (Danek & Wiley, 2017; Webb et al., 2016; Webb et al., 2019a). This evidence is consistent with the speculation that the unique variance explained by the Aha! ratings for solution memory in this and prior studies, over and above the variance explained by reaching a correct solution and the feeling of certainty, might be best represented as a benefit of pleasure or positive affect on memory.

In sum, these prior results suggest a role for both cognitive and affective components in the *insight memory advantage*, however this suggestion remains speculative until it is tested with a study that includes more direct measures of both affect and underlying insightful solution processes (restructuring or representational change). The main goal of the present research was to determine which aspects of insight problem solving are advantageous for recall of solutions. To foreshadow the results, the first study was able to replicate the *insight memory advantage* and its connection to both correctly generated solution attempts and subjective Aha! experiences as seen in Danek, Fraps, von Müller, Grothe, & Öllinger, 2013. The goal then for the second study was to implement more specific measures of both the affective component and the cognitive component to test both mechanisms more directly. These new measures are discussed following an overview of the methods used in both experiments.

### 1.1. Overview of methods for both experiments

The procedure generally followed that of Danek, Fraps, von Müller, Grothe, & Öllinger, 2013, 2014a, 2014b. The participants were undergraduate psychology students from the University of Illinois at Chicago who participated as part of an Introduction to Psychology subject pool and received course credit for their participation ($M_{age}$ = 19.04 years, $SD_{age}$ = 2.10, age range: 17–42 years, 71 males). The final sample size was 54 in the first study, and 127 in the second. The protocol was approved by the Institutional Review Board and the Office for the Protection of Research Subjects of the University of Illinois at Chicago, and all subjects gave written informed consent in accordance with the Declaration of Helsinki.

Participants individually viewed a set of 18 video clips of magic tricks[1] presented as a problem solving task ("Your task is to solve this puzzle and try to see through the magic trick."). Two practice trials preceded the set and the tricks were presented in randomized order. Professional magician Thomas Fraps (Abbott, 2005) performed the tricks (see https://www.youtube.com/watch?v=3B6ZxNROuNw for an example clip). The magic tricks took on average 17 s to view (with a range from 6 to 29 s) and participants generated a solution while the clip was playing, without any additional thinking time after the clip had ended. It was stressed that participants should only provide plausible solutions (no "magical" explanations like "a magic powd lets the coin disappear"), but that if they had an idea what the solution could be, then they should type it in even if they were not sure about it. They were told to press the space bar as soon as possible once they had a solution idea.

Before typing in their solution, participants were asked "Did you have an Aha! moment?". Participants marked their answer by shifting a slider to a point between "no" and "yes" on an unnumbered horizontal line.[2] Then they were asked to type in their solution and subsequently they rated how certain they were that their solution was correct, again selecting a point on an unnumbered sliding scale ranging from "uncertain" to "certain". Participants did not receive any feedback on the correctness of their solutions. This procedure was repeated for all 18 tricks. At the end of the first session, participants filled in a demographic data sheet. The first session lasted about 1 h.

After a one-week delay,[3] participants engaged in an unexpected recall task where they were tested on their memory for the solutions. They were instructed not to try to come up with any new solutions, only to remember the old ones from the week before. To refresh their memory, they saw a screen with two still shots as a reminder for that particular trick, one from the beginning and one from the end of the trick. They did not see the full videos again. They were prompted to type in the solution they remembered. After completing the procedure for all tricks which took half an hour or less, participants were thanked and debriefed.

[1] To increase solution rates and reduce the length of the experiment, we used only 18 of the original 34 tricks (Danek et al., 2013) with the highest solution rates from an earlier study with this population (Danek & Wiley, 2017).

[2] The previous study (Danek et al., 2013) used a dichotomous YES/NO response.

[3] The previous study (Danek et al., 2013) used a two-week delay.

Four key measures were available in both studies: Aha! and certainty ratings, correctness of solution attempts, and correctness of recall attempts. The rating measures were converted to numerical values using a continuous scale of 0–100 units to represent the length of the lines. The location of the point was measured from the left end (so that placement at the very left represented 0, the very right represented 100, and the middle represented 50) for all rating responses in Experiments 1 and 2.

Another key measure was the correctness of the solution attempts. Using a coding manual (compiled with the help of the magician), participants' solutions were coded as correct (methods that the magician actually used or alternative methods verified as plausible) or incorrect (partial solutions, implausible methods, or impossible solutions with respect to the conditions seen in the video clip) by two independent raters. Calculated across both datasets together, the two-way random intraclass correlation coefficient (absolute agreement), ICC (2,2) was 0.84 with a 95% confidence interval of [0.83;0.85] indicating a good level of agreement according to the conventions set out in Koo and Li (2016). For all coding measures, conflicting cases were resolved by a third rater.

The final key measure collected in both studies was memory for the solution attempts after one-week delay. Two independent raters coded responses into two categories: recalled and not recalled. The first category was assigned if participants provided solutions identical to those which they had given or were given previously. The second category comprised completely forgotten solutions as well as incorrectly remembered ones (solutions which were not identical to those previously given). Across both datasets, the two-way random intraclass correlation coefficient (absolute agreement), ICC (2, 2) was 0.79 with a 95% confidence interval of [0.77;0.80] indicating a good level of agreement.

### 1.2. New measures introduced in experiment 2

Experiment 2 added new measures to assess the role of more specific affective and cognitive components in the *insight memory advantage*.

#### 1.2.1. Measuring feelings associated with the Aha! experience

The Aha! experience is a multidimensional construct consisting of feelings of pleasure, relief, suddenness, and certainty/confidence (Danek et al., 2014a; Danek & Wiley, 2017; Webb et al., 2016, 2017). Roles for surprise and impasse have also been suggested, but are less clear. To tease apart the contribution of these different dimensions, past work has taken the approach of breaking down the Aha! experience into separate components and assessing them individually. Danek et al. (2014a) used two retrospective measures, open-ended descriptions of the Aha! experience and ratings on 5 dimensions (pleasure, certainty, suddenness, surprise, impasse). The feeling of pleasure featured prominently both in the open-ended descriptions, as well as in the ratings where it was more strongly endorsed than any of the other dimensions. Impasse was the least strongly endorsed dimension, with the remaining dimensions not significantly differing from each other. Two other dimensions (relief and drive/motivation to continue solving) emerged from the open-ended descriptions. Danek & Wiley, 2017 then collected ratings on 6 dimensions (pleasure, certainty, suddenness, surprise, relief, drive) immediately after each solution attempt. Using regression to identify the individual contribution of each dimension to the overall feeling of Aha!, the dimensions pleasure, suddenness, certainty and relief were found to be unique predictors of Aha! for correct solutions. For incorrect solutions, surprise instead of relief contributed unique variance. Taken together, both prior studies suggest a primary role for pleasure in relation to the Aha! experience. Following Danek & Wiley, 2017, the approach used in Experiment 2 was to ask participants to rate 6 separate dimensions of the Aha! experience following each trick so that these feelings could be used as predictors of solution memory.

#### 1.2.2. Measuring representational change and sudden restructuring

The insightful solution process has been posited to involve a sudden change or discontinuity that is thought to represent the sudden restructuring of a problem representation (Metcalfe & Wiebe, 1987). This may be more likely to occur during the solution of some problems than others, as shown by Metcalfe and Wiebe (1987), and may be especially likely on problems that cue an incorrect initial representation such as magic tricks, and thus require representational change before solution.

It is notoriously difficult to assess whether the hypothetical mechanism of representational change has actually occurred as this requires some measure of the solvers' internal problem representations (Ash et al., 2009). One method that has been used to gain access to these internal representations is to have solvers repeatedly rate how important for solution (or how related with each other) individual problem elements seem to be. In Experiment 2 of their pioneering study, Durso et al. (1994) plotted similarity ratings over time and found evidence for incremental changes in the problem representation toward a correct solution during the solution of a verbal puzzle. Using a parallel approach, Cushen and Wiley (2012) were able to track changes in individual solvers' problem representations by asking for repeated importance-to-solution ratings for problem elements over time. Increases in perceived importance between initial and final ratings for critical problem elements showed individuals changing toward a correct problem representation. In addition, this method was also able to identify both incremental and sudden change patterns for individuals.

Recently, Danek, Williams, & Wiley, 2020 combined this method with the approach of obtaining subjective Aha! ratings after each solution to test for a relationship between the cognitive and affective components of insightful problem solving. To test for sudden restructuring while solving magic tricks, instead of rating problem elements, participants were asked to repeatedly rate 6 action verbs for their “importance to solution”. For each trick, one verb described the correct solution, one verb described a false solution, and 4 other verbs were unrelated distractors. For example, in the “Match through Match” trick, the verb that represented the false solution was “breaking” while the target verb which represented the correct solution was “raising”. This study found that sudden changes in ratings toward a correct representation were more likely to elicit an Aha! experience than were incremental changes, offering first empirical support for the theoretically assumed relationship between cognitive and affective components of insight problem solving. By adding this verb rating method to Experiment 2, the impact of representational change or sudden restructuring as cognitive component underlying the *insight memory advantage* could be tested.

## 2. Experiment 1

The purpose of Experiment 1 was to replicate the *insight memory advantage*. Following Danek, Fraps, von Müller, Grothe, & Öllinger, 2013, the main hypothesis that was tested in this study was that both Aha! experiences and correctly generated solutions would predict better memory for solutions on a set of magic tricks. Following Danek et al. (2014b), it was expected that when participants generate correct solutions it should lead to stronger Aha! experiences than when participants generate incorrect solutions. As pointed out by Danek & Salvi, 2020, such an *accuracy effect* on insight ratings has now been established in several studies (for magic tricks by Danek et al., 2014b; Danek & Wiley, 2017; Danek, Williams, & Wiley, 2020; Hedne et al., 2016; but also for other tasks such as rebus puzzles or anagrams, as shown by Salvi et al., 2016; Threadgold et al., 2018; Webb et al., 2016, 2017; Webb et al., 2019a). Based on Danek et al. (2014b), it was also expected that solutions associated with stronger Aha! experiences should also be rated with higher certainty, that solutions emerging later in the solution process would be less likely to evoke strong Aha! experiences, and that solutions associated with stronger Aha! experiences should also be rated with higher certainty. Given the expected relations among feelings of certainty, feelings of Aha!, and generating correct solutions, analyses predicting solution recall also explored interactions among these factors, to better understand their roles in the *insight memory advantage*.

### 2.1. Method

This study involved a few specifications to the general methods described above. First, at the start of the study participants were instructed to base their rating on the following description of what an Aha! moment typically feels like (back-translated with minor modifications from the German instruction of Danek, Fraps, von Müller, Grothe, & Öllinger, 2013; which had been originally adapted from Jung-Beeman et al., 2004):

> "An Aha! moment is when the solution dawns on you suddenly and everything is clear immediately, in a flash. You are relatively confident that your solution is correct. In contrast, if the solution occurs to you slowly and in steps that would not be an Aha! moment. As an example, imagine a light bulb that is switched on all at once in contrast to slowly turning up the lights. Have you ever experienced an Aha! moment, perhaps during studying? For each solution, we ask for your subjective rating whether it felt like an Aha! moment or not. There is no right or wrong answer. Just follow your intuition."

Second, as shown in Fig. 1, in this study each video clip could be viewed up to a maximum of three times. Pressing the spacebar ended the video and advanced participants to the Aha! rating screen. If the spacebar was not pressed (because participants could not come up with a solution idea), then following the third viewing of the video, the solution to the trick was shown.[4] This was in order to keep the memory load the same for each participant independent of how many tricks they solved by themselves. No Aha! ratings or certainty ratings were collected on these "shown solution" trials.

Third, in this study participants returned to the lab for a second session and recall for solutions was assessed in a two-step process. First participants indicated if they remembered having seen a trick (based on the screenshots), and then only if they said "YES" to the first question, were they prompted to recall the solution from the prior week.

### 2.2. Results

#### 2.2.1. Descriptive statistics

In total, 54 participants being presented with 18 tricks yielded 972 observations. There were 51 missing values due to technical errors (skipped trials), leaving 921 valid trials. Of those, no solution was generated on 19.9% of the trials (183 observations) and participants were shown solutions. On those trials where participants had generated a solution (738 observations), 49% (362) were correct solutions, and 51% (376) were incorrect solutions. Failures to recall the trick entirely were collapsed into failures to recall the solution. Initial solutions were correctly recalled one week later on 46% of tricks where a solution was generated. Splitting by solution correctness showed that of all correct solutions, 57.7% were correctly recalled, and of all incorrect solutions, 35.4% were correctly recalled.

Because the correct solution was presented at the end of every trial where no solution was generated, the effect of generating versus being shown correct solutions on solution recall could be tested. However, because participants did not also complete Aha! ratings on these trials, shown-solution trials could not be included in most analyses. The difference in solution recall due to generating or being shown the correct solution is considered in a final section.

#### 2.2.2. Relations with Aha! experiences

Before testing the main question, simple correlations among predictor variables were tested and are shown in Table 1.[5]

To provide a replication of the *accuracy effect* established in Danek et al. (2014b), a mixed effects model (number of observations = 738) entering solution correctness as a fixed effect, and fitting random intercepts for subjects ($Z = 4.10$, $p < .001$) and tricks ($Z = 1.95$, $p = .05$), was calculated to test whether correct and incorrect solutions would be associated with differences in average Aha! ratings.[6] Correct solutions had higher Aha! ratings ($M = 65.66$, $SD = 28.42$) than incorrect solutions ($M = 51.28$, $SD = 30.32$), $F(1, 736) = 50.17$, $p < .001$, $d = 0.52$.[7] The same significant relation was seen if tested as a mixed effects binary logistic regression with Aha! ratings predicting whether tricks were solved correctly. As can be seen in Table 1, certainty ratings were also correlated with correct solutions, and with Aha! ratings. The correlation between Aha! ratings and recall also provides a replication of the *insight memory advantage* ($d = 0.51$).

#### 2.2.3. Do Aha! experiences, solution correctness, and feelings of certainty predict recall of solutions?

To test the main question of whether the strength of the Aha! experience, solution correctness, and feelings of certainty were each independently related to memory for the generated solutions, a mixed effects model (number of observations = 738) was used to perform a binary logistic regression on solution recall including all three predictors (Aha! experience, solution correctness, and certainty) as fixed effects. Random intercepts were included for both participants ($Z = 2.25$, $p = .02$) and tricks ($Z = 2.10$, $p = .04$). The overall model was significant, $F(3, 734) = 24.02$, $p < .001$. Significant unique variance was contributed by all three predictors, suggesting that not just correct solution, but the feeling that one had reached a correct solution (i.e., certainty), and also the feeling of Aha! all independently predicted solution memory, as shown in Table 2. Neither the two-way interaction between solution correctness and Aha! ($p = .64$) nor solution correctness and certainty ($p = .14$) nor Aha!

[4] In contrast to Danek et al. (2013, 2014a, 2014b) where solutions to unsolved tricks were given all at once at the end of the experiment, now solutions were shown immediately after each unsolved trick.

[5] For all correlational analyses, values are Pearson coefficients if both variables are continuous, point-biserial coefficients if one variable is binary, or phi coefficients if both variables are binary.

[6] For all mixed effects analyses in this paper, mixed models were computed using the generalized linear mixed model (GENLINMIXED) procedure from IBM SPSS Statistics v. 25 with participants and items (tricks) as random effects (Baayen et al., 2008). Continuous predictor variables were grand-mean centered before being entered into regression models. Dummy-coding (1 for correct solution, 0 for incorrect) was used to compute interaction terms between continuous predictors and solution correctness.

[7] As an estimate of effect size, Cohen's $d$ was calculated using $t$-to-$d$ transformation with $d = \frac{2t}{\sqrt{df}}$ (Rosenthal & Rubin, 2003), as recommended by Page-Gould (2013).

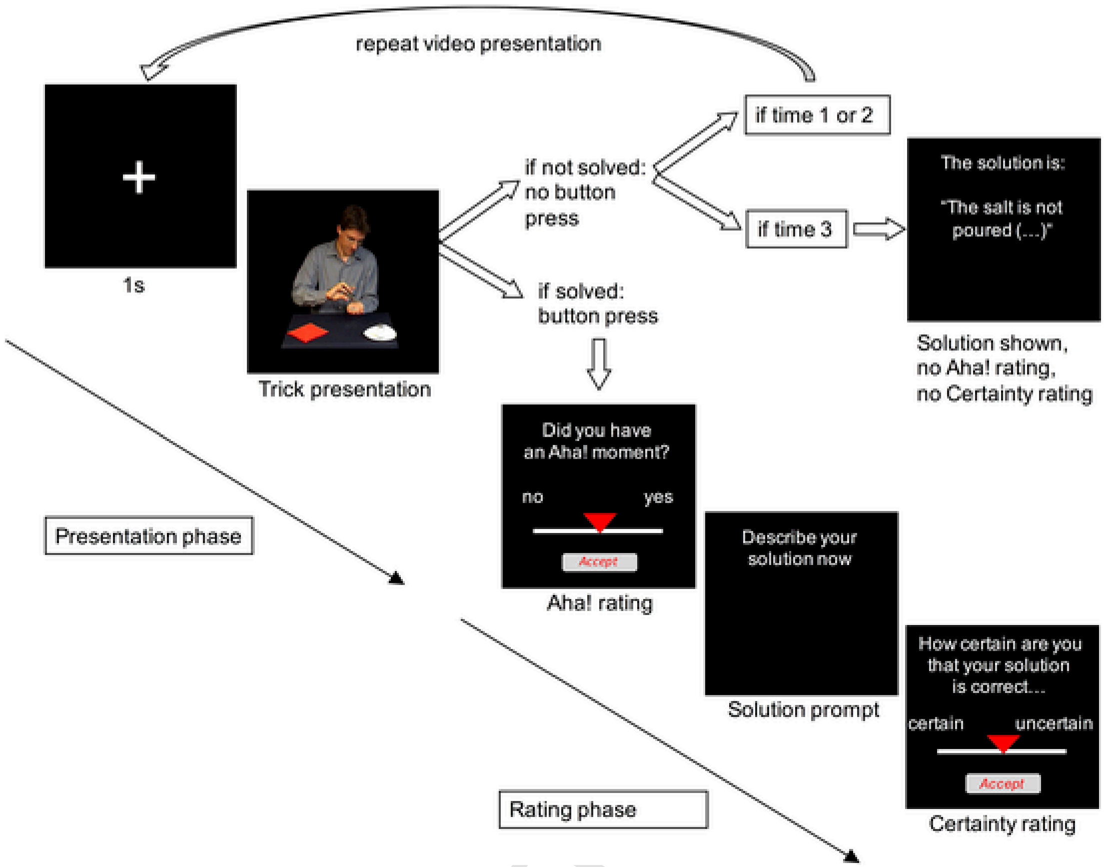


**Fig. 1.** Sequence of one trial during the first session in Experiment 1.

**Table 1**
Simple correlations among predictor variables and solution recall in Experiment 1.

| | Aha! | Certainty | Correctness | Recall |
|---|---|---|---|---|
| Aha! | – | 0.58 ** | 0.24 ** | 0.25 ** |
| Certainty | | – | 0.32 ** | 0.28 ** |
| Correctness | | | – | 0.22 ** |

*Note.* N = 738.
** $p < .01$.

**Table 2**
Binary logistic regression model (mixed effects) with Aha!, correctness, and certainty as predictors of solution recall in Experiment 1.

| | Unstandardized coefficient B | SE B | t | p | d |
|---|---|---|---|---|---|
| Constant | −0.467 | 1.382 | | | |
| Aha! | 0.010 | 0.003 | 2.90 | $p = .004$ ** | 0.21 |
| Correctness | 0.598 | 0.189 | 3.17 | $p = .002$ ** | 0.23 |
| Certainty | 0.018 | 0.004 | 4.16 | $p < .001$ ** | 0.31 |

*Note.* N = 738.
** $p < .01$.

and certainty ($p = .33$) reached significance when added to the model.

*2.2.4. Does generating versus being shown a correct solution improve solution memory?*

A supplementary analysis was performed to test whether a *generation effect* might be seen in this study, following Kizilirmak et al. (2016). A mixed effects model entering solution type (generated correct vs. shown correct, 545 observations) as a fixed effect, and fitting random intercepts for subjects ($Z = 1.54$, $p = .12$) and tricks ($Z = 2.19$, $p = .03$), was used to perform a binary logistic regression testing whether solution recall would be improved by having generated a correct solution over having been shown the correct solution. In contrast to earlier work showing advantages of generating a correct solution on memory (Kizilirmak et al., 2016), better recall was seen for shown solutions in this study (Shown $M = 0.68$, $SD = 0.47$; Correctly Generated $M = 0.58$, $SD = 0.50$, $F(1, 543) = 4.58$, $p = .03$, $d = 0.18$). It is possible that showing the correct solution only after three failed attempts may have been more like the delayed revelation condition that improved memory for solutions in Auble et al. (1979).

Further inspection of the data suggested that this procedure may have had negative repercussions on the motivation to generate a solution at all. The number of unsolved problems increased as the study progressed and participants may have stopped responding in order to see the correct solution. To test this possibility, trial number (out of 18) was entered as fixed effect into a mixed effects model (binary logistic regression) predicting solution type (generated correct, shown correct, 545 observations). Random intercepts were again included for subjects ($Z = 0.87$, $p = .39$) and tricks ($Z = 2.39$, $p = .02$). The overall model was significant, $F(1, 543) = 4.63$, $p = .03$. The number of shown solutions were significantly predicted by trial number, and became more frequent over time ($B = 0.05$, $t = 2.15$, $p = .03$). Another possible reason for the improved memory for these revealed solutions could be that students who were most curious to learn the correct solution for a trick withheld their responses, and their interest made these solutions highly memorable.

These issues suggested it would be better to not provide solutions in Experiment 2.

#### 2.2.5. Relation between number of trick presentations and Aha! experiences

A second supplementary analysis tested whether solutions emerging later in the solution process would be less likely to evoke strong Aha! experiences, following Danek et al. (2014b). Tricks were presented up to three times and a solution could be attempted during the first, second or third presentation. The frequency of solution attempts (correct or incorrect) was not evenly spread across the three presentations. Few solutions were generated following the first presentation (13%). The majority of solutions were generated after the second presentation (54.6%) and the remaining ones (32.4%) after the third.

To test whether Aha! ratings might decrease after multiple presentations of a trick, a mixed effects model was calculated (number of observations = 738), entering number of presentations as fixed effect and fitting random intercepts for subjects ($Z = 4.02$, $p < .001$) and tricks ($Z = 2.17$, $p = .03$). A significant effect was found for number of presentations, $F(1, 735) = 10.58$, $p < .001$. Pairwise comparisons showed that solutions generated after the third presentation of the trick were associated with lower Aha! ratings ($M = 52.11$, $SD = 31.16$) than solutions generated after the first ($M = 57.17$, $SD = 32.22$) or second presentation ($M = 62.30$, $SD = 28.61$). These results are consistent with Danek et al. (2014b) and their implications will be further discussed in interpreting results of Experiment 2.

### 2.3. Discussion

Problem solvers gave higher Aha! ratings to correct solutions than to incorrect solutions which provides yet another example for the *accuracy effect* on insight ratings (Danek et al., 2014b). Further, Aha! ratings strongly correlated with certainty ratings (even though solvers received no feedback), and solutions that were generated after spending an extended time in a solution attempt (i.e., after the third exposure to the trick) were associated with weaker Aha! experiences, also replicating prior findings.

With regard to the *insight memory advantage*, solutions associated with Aha! experiences were remembered better than those not associated with Aha! experiences. Further, not only the Aha! experience, but also correctness of solution predicted better solution memory, replicating the results of Danek, Fraps, von Müller, Grothe, & Öllinger, 2013 and Kizilirmak et al. (2016). Thus, the results of Experiment 1 follow the results of prior studies showing an independent contribution of both correct solutions and the Aha! experience on memory for solutions. The prevailing explanation for this pattern of results is that while correct solutions may be capturing variance coming from the cognitive component of the solution process, that the subjective Aha! ratings may be capturing affective components.

Independent effects were found for both feelings of certainty and feelings of Aha! over and above the effect of generating a correct solution. This analysis showed that not just actually reaching a correct solution, but also having the *feeling* that one had reached a correct solution, was independently predictive of later solution memory. Further, there remained unique variance that was still explained by the Aha! rating. Based on evidence suggesting a primary role for positive affect in Aha! ratings (Danek et al., 2014a; Danek & Wiley, 2017; Kizilirmak et al., 2016), positive affect or pleasure seems a strong candidate to explain the independent contribution of the Aha! ratings even once correct solution and feelings of certainty are accounted for. Thus, one goal for Experiment 2 was adding measures of distinct feelings that may underlie the Aha! experience (especially pleasure) together with more specific measures of representational change and sudden restructuring, so that their role in the memory advantage could be tested.

## 3. Experiment 2

The main question that was tested in Experiment 2 was which of the more specific measures of feelings associated with Aha! experiences and which representational change indices would predict memory for the generated solutions. If the pleasurable affective experience associated with Aha! moments is responsible for the *insight memory advantage*, then recall of solutions should increase with pleasure ratings. If the feeling of certainty associated with Aha! moments is responsible for the *insight memory advantage*, then recall of solutions should increase with certainty ratings. Further, if the memory advantage is the result of insightful solution processes, then recall for solutions should not only be higher when a correct solution is generated, but also when there is evidence of representational change. A particular boost should be seen when a problem is solved via sudden restructuring than via other patterns (e.g., incremental ones).

### 3.1. Method

Methods largely followed the same procedures as in Experiment 1, but two new sets of measures were added (see Fig. 2).

#### 3.1.1. Ratings for 6 individual dimensions of the solution experience

The certainty rating scale was identical to the one used in Experiment 1, but 5 more rating scales were added (the same ones as used in Danek & Wiley, 2017). Participants rated their subjective solution experiences after each trick using unnumbered scales with the following wording:

1. Pleasure: “At the moment of solution, my feelings were… (unpleasant - pleasant)”
2. Surprise: “The moment of solution was… (not surprising - surprising)”
3. Suddenness: “This solution came to me… (in steps - all at once)”
4. Relief: “At the moment of solution, I felt… (tense - relieved)”
5. Certainty: “How certain are you that your solution is correct: (uncertain - certain)”
6. Drive: “I am looking forward to the next trick… (no - yes)”

The dimension ratings started with 4 ratings (pleasure, surprise, suddenness and relief), then participants were prompted to type in their solution (or “no idea”) and finished the trial with 2 more ratings (certainty and drive). The direction of the scales was reversed for three of the ratings, and counterbalanced across subjects. One half of the participants saw the left-right anchors for the pleasure, suddenness and drive scales with the remaining anchors reversed. The other half of the participants had the surprise, relief and certainty anchors preserved and the others reversed. The position of the Aha! rating was also counterbalanced. It either came before or after all of the dimension ratings.

#### 3.1.2. The verb rating task

To provide a measure of each solver's problem representation, participants were asked to rate a set of 5 verbs[8] after each trick pre-

[8] Ratings obtained in Danek et al. (2020) were used to identify target verbs that received a low rating (defined as <60% of the scale) by correct solvers, or target verbs that received a high rating (> 60%) by non-solvers. Two research assistants generated suggestions for new verbs, first separately, then together with one of the authors (AD). Of the 18 original target verbs, 5 were replaced.

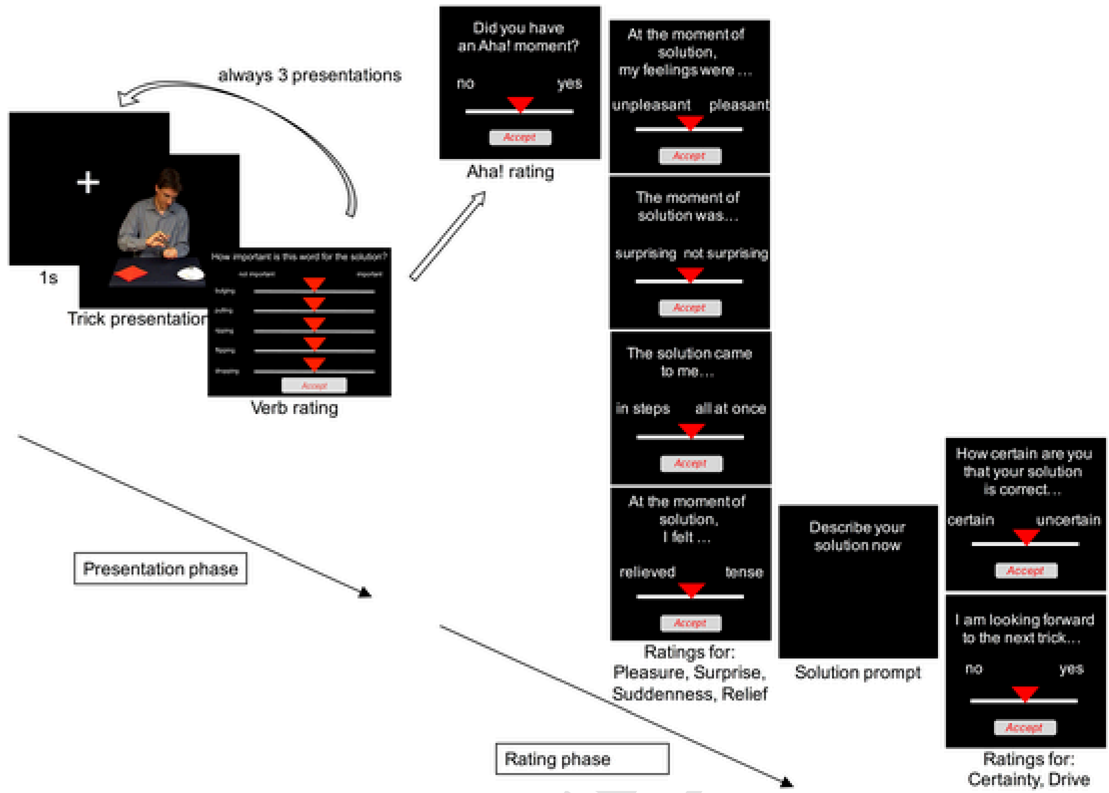


**Fig. 2.** Sequence of one trial during the first session of Experiment 2.

sentation with regard to how well they described the solution (“How important is this word for the solution?”), see third screen “Verb rating” in Fig. 2 (presentation phase). This was meant to capture whether the verb matched a solution that was currently being considered by the solver (a situational judgment, not a metacognitive judgment, as distinguished by Ash & Wiley, 2008). During practice, they were told to look at each verb individually and make a rating for it by selecting a point on an unnumbered line that was anchored by “not important” on the left and “important” on the right. For each trick, there was one “target verb” that corresponded to the correct solution, one “biased verb” consistent with the biased initial representation of the problem that the magician attempted to promote, and three distractor verbs.[9] Verb order was varied across tricks and participants, to avoid, for example, the target verb always being first. The instruction was: “After each viewing, we will ask you to rate how important some words are for the solution of the magic trick, if they describe the solution or not. This is not about the magic trick itself, only about the solution.”

### 3.1.3. Additional changes

There were a few other changes in procedures. First, the Aha! rating scale was the same as in Experiment 1, but the description of the Aha! experience was omitted in order to avoid the ambiguity caused by giving Aha! prompts which contain several different dimensions (as discussed in Cushen & Wiley, 2012; Danek, 2018; Grunewald & Bowden, 2018). Participants were simply prompted to rate the extent to which they felt an Aha! experience. A pilot study indicated that participants were able to rate their Aha! experience without any additional instruction. To confirm that all participants understood what an Aha! moment was, at the end of the first session they were asked “Do you know what is meant by Aha! moment?”. After answering with either yes or no, they were asked “Please describe what you mean by an Aha! moment.” The first question was answered with “no” by only 4 participants. There were 5 more participants who answered “yes”, but nevertheless defined Aha! in an idiosyncratic way (e.g., “like wow it was a crazy trick” or “when an audience gets surprised by watching something in joy”). These 9 participants were removed from analyses and not included in the final sample. The survey also asked participants whether they had been familiar with any of the magic tricks prior to the study so these instances could be excluded, too.

In contrast to Experiment 1, participants were not shown correct solutions for any of the tricks, and participants had to go through a fixed sequence of three viewings and three verb ratings before they could provide a solution. This was in order to obtain verb rating data at all three timepoints. After completing the third verb rating, the solution experience for each trick was assessed using the Aha! and dimension ratings as described above.

The procedure for the second session was very similar to Experiment 1 except that it was conducted online instead of in the lab. After exactly 1 week, participants received a link to an online survey

[9] A new category (“biased” verb) was developed to replace the “false solution” verb used in Danek et al. (2020) which represented common incorrect solutions. Two assistants watched the video clips repeatedly to identify verbs that matched the initial, wrong problem representation that was subtly suggested by the magician. They considered which wrong assumptions a naïve observer would make, chose the main assumption acting as a constraint and preventing a solution, and described it in one verb. After conducting this process independently, the assistants established the final biased verb list in discussion with AD. For the example trick described previously, the false solution verb “breaking” was replaced with the biased verb “touching” because the trick presentation leads the observer assume that the matches touch each other (which in reality, they do not). The 3 distractor verbs were taken from the list of target and biased verbs from the original set of 34 tricks. The number of distractors was reduced from 4 to 3 to avoid repeating too many verbs.

which they had to take on the same day. They saw the same instructions and the same still shots as in Experiment 1, and were asked to type in their recalled solutions to each trick. In contrast to Experiment 1, the initial question about whether they remembered the trick or not was omitted since in Experiment 1 this differentiation was not used any further, but collapsed into the "not recalled" category.

### *3.2. Results*

#### *3.2.1. Descriptive statistics*

In total, 127 participants being presented with 18 tricks yielded 2286 observations. Due to participants reporting to be familiar with the solution a priori, 16 observations were discarded, leaving 2270 observations. Of those, no solutions were generated for 331 trials and thus those observations could not be analyzed. On 76 trials, solvers provided plausible alternative solutions that differed from the methods that were actually used by the magician. Because the target verb had been developed specifically to correspond to the actual method, trials with plausible alternative solutions were not used. Analyses were based on the remaining 1863 observations. Of these, 50.1% (933 observations) were correctly solved, and 49.9% (930 observations) were incorrectly solved, which is similar to the rates in Experiment 1, suggesting that verb ratings did not act as cues that altered solution rates.

Recall of initial solutions was correct on 70% of tricks where solutions were generated. This rate is somewhat higher than the recall rate seen in Experiment 1, probably due to the fact that all participants watched each trick three times. In addition, this higher recall rate suggests that completing the second session online in Experiment 2 did not seem to have a negative impact.

#### *3.2.2. Initial analyses to test for replications from experiment 1*

*3.2.2.1. Relation between solution correctness and Aha! experiences* The simple correlations among predictor variables are shown in Table 3. As in Experiment 1, a mixed effects model (number of observations = 1863) entering solution correctness as a fixed effect, and fitting random intercepts for both participants ($Z = 6.86$, $p < .001$) and tricks ($Z = 2.54$, $p = .01$), was calculated to test whether correct and incorrect solutions would be associated with differences in average Aha! ratings. Again, there was an *accuracy effect* as correct solutions had higher Aha! ratings ($M = 60.50$, $SD = 32.4$) than incorrect solutions ($M = 45.55$, $SD = 30.92$), $F(1, 1861) = 84.08$, $p < .001$, $d = 0.43$. The same significant relation was seen if tested as a mixed effects binary logistic regression with Aha! ratings predicting whether tricks were solved correctly. As can be seen in Table 3, certainty ratings were again correlated with correct solutions, and with Aha! ratings. And again, the correlation between Aha! ratings and recall replicated the *insight memory advantage* ($d = 0.30$).

**Table 3**
Simple correlations among predictor variables and solution recall in Experiment 2.

| | Aha! | Certainty | Correctness | Recall |
|---|---|---|---|---|
| Aha! | _ | 0.45 ** | 0.23 ** | 0.15 ** |
| Certainty | | _ | 0.34 ** | 0.27 ** |
| Correctness | | | _ | 0.23 ** |

*Note.* $N = 1863$.
** $p < .01$.

*3.2.2.2. Do Aha! experiences, solution correctness, and feelings of certainty predict recall of solutions?* To test the question of whether the strength of the Aha! experience, solution correctness, and feelings of certainty were independently related to memory for the generated solutions, a mixed effects model (number of observations = 1863) was used to perform a binary logistic regression on solution recall as in Experiment 1. Random intercepts were included for both participants ($Z = 4.65$, $p < .001$) and tricks ($Z = 2.24$, $p = .025$). Aha! ratings, solution correctness, and certainty were entered as fixed effects. The overall model was significant, $F(3, 1859) = 43.51$, $p < .001$. As shown in Table 4, solution recall was again predicted by Aha! ratings, solution correctness and certainty, with all three of them contributing significant unique variance. As in Experiment 1, neither the interaction between solution correctness and Aha! ratings ($p = .43$), nor the interaction between solution correctness and certainty ratings ($p = .20$), nor the interaction between Aha! and certainty ($p = .75$) were significant when added to the model.

#### *3.2.3. Analyses on individual feeling dimensions and the Aha! experience*

These analyses start by first showing the relation of ratings on each individual feeling dimension to the overall Aha! rating. The second analysis tests whether pleasure and certainty are responsible for better solution memory.

*3.2.3.1. Which feeling dimensions predict Aha?* The simple correlations in Table 5 show that all 6 dimensions were significantly related to the Aha! rating (collapsed across correct and incorrect solutions). This supports the idea that the Aha! experience is a multidimensional construct (Danek & Wiley, 2017).To understand the unique relation of each dimension to Aha!, a mixed effects model (number of observations = 1863) was calculated to predict the Aha! ratings, including each dimension as fixed effect. Random intercepts were included for both participants ($Z = 6.68$, $p < .001$) and tricks ($Z = 1.81$, $p = .07$). The overall model was significant, $F(6, 1856) = 140.87$, $p < .001$. As shown in Table 6, all 6 dimensions were found to be unique predictors of the Aha! experience during problem solving attempts (collapsed across correct and incorrect solutions).

*3.2.3.2. Which feeling dimensions predict recall of solutions?* To understand the unique relation of each dimension to solution recall, a mixed effects model (number of observations = 1863) was calculated to predict recall, including each dimension as fixed effect. Random intercepts were included for both participants ($Z = 4.69$, $p < .001$) and tricks ($Z = 2.17$, $p = .03$). The overall model was significant, $F(6, 1856) = 17.63$, $p < .001$. As shown in Table 7, only pleasure and certainty were found to be unique predictors of solution recall.

**Table 4**
Binary logistic regression model (mixed effects) with Aha!, correctness, and certainty as predictors of solution recall in Experiment 2.

| | Unstandardized coefficient B | SE B | t | p | d |
|---|---|---|---|---|---|
| Constant | 0.615 | 1.268 | | | |
| Aha! | 0.005 | 0.002 | 2.23 | $p = .026$ * | 0.10 |
| Correctness | 0.774 | 0.132 | 5.86 | $p < .001$ ** | 0.27 |
| Certainty | 0.015 | 0.002 | 6.25 | $p < .001$ ** | 0.29 |

*Note.* $N = 1863$.
* $p < .05$.
** $p < .01$.

**Table 5**
Simple correlations between individual dimensions and Aha! ratings in Experiment 2.

| Dimension | Aha! | Pleasure | Surprise | Suddenness | Relief | Certainty | Drive |
|---|---|---|---|---|---|---|---|
| Aha! | _ | 0.52 ** | 0.05 * | 0.07 ** | 0.43 ** | 0.45 ** | 0.24 ** |
| Pleasure | | _ | 0.02 | 0.16 ** | 0.58 ** | 0.47 ** | 0.27 ** |
| Surprise | | | _ | −0.41 ** | 0.00 | −0.27 ** | 0.19 ** |
| Suddenness | | | | _ | 0.15 ** | 0.34 ** | −0.04 |
| Relief | | | | | _ | 0.45 ** | 0.22 ** |
| Certainty | | | | | | _ | 0.14 ** |

*Note.* N = 1863.
* $p < .05$.
** $p < .01$.

**Table 6**
Linear regression model (mixed effects) of dimensions as predictors of the Aha! experience in Experiment 2.

| | Unstandardized coefficient B | SE B | t | p |
|---|---|---|---|---|
| Constant | 53.06 | 15.633 | | |
| Pleasure | 0.286 | 0.031 | 9.08 | $p < .001$ ** |
| Surprise | 0.109 | 0.025 | 4.35 | $p < .001$ ** |
| Suddenness | −0.054 | 0.020 | −2.67 | $p = .008$ ** |
| Relief | 0.201 | 0.032 | 6.33 | $p < .001$ ** |
| Certainty | 0.333 | 0.025 | 13.31 | $p < .001$ ** |
| Drive | 0.077 | 0.027 | 2.89 | $p = .004$ ** |

*Note.* N = 1863.
** $p < .01$.

**Table 7**
Binary logistic regression model (mixed effects) with all 6 dimensions as predictors of solution recall in Experiment 2.

| | Unstandardized coefficient B | SE B | t | *p* |
|---|---|---|---|---|
| Constant | 0.985 | 1.278 | | |
| Pleasure | 0.008 | 0.003 | 2.39 | $p = .017$ * |
| Surprise | −0.005 | 0.003 | −1.68 | $p = .093$ |
| Suddenness | 0.003 | 0.002 | 1.52 | $p = .129$ |
| Relief | −0.002 | 0.003 | −0.56 | $p = .578$ |
| Certainty | 0.016 | 0.003 | 6.41 | $p < .001$ ** |
| Drive | 0.003 | 0.002 | 1.08 | $p = .282$ |

*Note.* N = 1863.
* $p < .05$.
** $p < .01$.

#### *3.2.4. Analyses on measures of representational change*

Three indices of representational change were derived from the importance-to-solution verb ratings collected in this study. One measure focused on the timing of the change and specifically categorized whether the target verb ratings showed evidence of *sudden restructuring*. This was done by considering the patterns of target verb ratings over time. A second, more general measure was the magnitude of overall *change toward a correct problem representation* which was derived by considering the amount of change from initial to final target verb ratings. As a complement to the second measure, the third measure was the magnitude of overall *change away from an incorrect problem representation* derived from the initial to final biased verb ratings. The coding and computations that allowed for the derivations of each of these measures are described before each corresponding analysis.

*3.2.4.1. Coding for target verb rating patterns* The procedure for identifying patterns over time in the importance-to-solution ratings for the target verbs was based on the Danek, Williams, & Wiley, 2020. In the prior study, not only incremental and sudden increase patterns were found, but also decreasing patterns (indicating that target verbs were rated as less important as problem solving proceeded) and flat patterns (verbs rated consistently across all ratings). Remaining patterns were coded as 'other' and not included. One extension made to the categories originally used by Danek, Williams, & Wiley, 2020 was the addition of a separate "high flat" category for flat patterns that started with a high initial rating on the target verb already after the first exposure (which may represent tricks that are either solved very quickly or with no initial misrepresentation) versus all other flat patterns.[10] A second extension was separating the sudden increases into those that occurred between the first and second exposure from those that occurred between the second and third exposure.[11] This resulted in 6 distinct categories of patterns that were seen across the three target verb rating timepoints (decreasing, flat (not high), incremental increase, sudden increase before third rating, sudden increase before second rating, and high flat). Fig. 3 shows average graphs illustrating each of the 6 patterns.To aid the coding process, a line graph was created from each participant's target verb ratings across the three timepoints for each individual trick (similar to the graphs in Fig. 3). As in prior studies (Cushen & Wiley, 2012; Danek, Williams, & Wiley, 2020), two

[10] A re-examination of the Danek et al. (2020) results revealed that many flat patterns (57%) already had very high ratings on target verbs after watching the trick for the first time. Throughout further viewings, ratings remained high, indicating that participants had probably found the solution right away. Of the flat patterns that were associated with correct solutions, 73% were "high flat" patterns, compared to only 43% of flat patterns associated with incorrect solutions. Although this "high flat" pattern category was not analyzed separately in the previous study, it is possible that some participants may have gotten past the incorrect representation very suddenly (too quickly to be captured by multiple ratings), while others may have started with a correct initial problem representation, eliminating the need for restructuring. Either of these cases would make this category of "high flat" responses distinctly different from other flat patterns, which presumably reflect a lack of change from an incorrect problem representation during solution attempts. Other work has also suggested the need to consider immediate solutions separately. For example, a think-aloud study Cranford and Moss (2012) found that immediate solutions, although rated as insights by participants, exhibited hardly any characteristics of the insight problem solving process such as impasse or restructuring. They concluded that immediate solutions result from a different process more similar to memory retrieval.

[11] Re-examination of the Danek et al. (2020) data showed that the majority of sudden increases did occur in the earlier time window (74.5%). When Aha! ratings associated with the two time windows were examined separately, Aha! ratings were lower when sudden increases occurred between the second and third viewing ($M = 3.18$, $SD = 1.95$) than when they occurred between the first and second viewing ($M = 3.57$, $SD = 1.68$). This finding is further corroborated by the general decrease in Aha! ratings for solutions that occurred during the third viewing in Experiment 1. For these reasons, the sudden increase category was split by time window in Experiment 2.

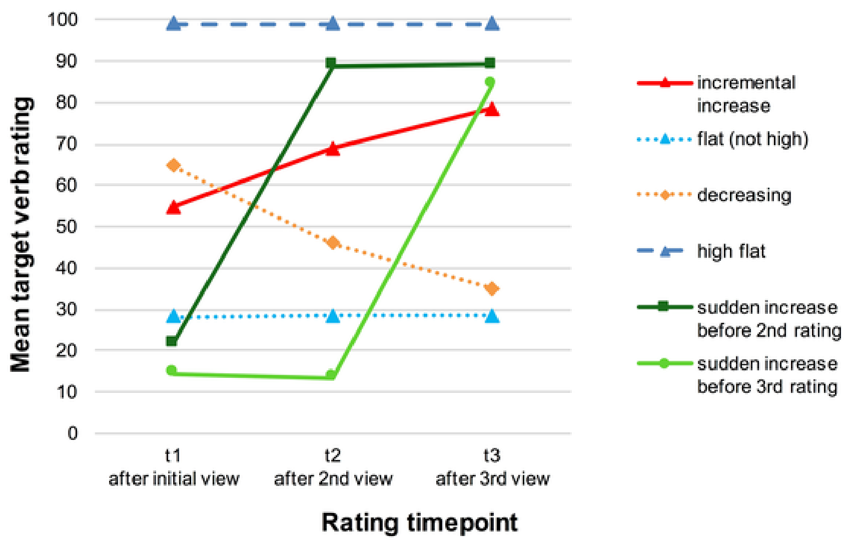


**Fig. 3.** Mean target verb ratings (on a scale from 0 to 100) in Experiment 2 across the three rating timepoints, as a function of pattern, collapsed across correct and incorrect solutions.

independent raters conducted a visual analysis of all 2286 graphs and made a category judgment for each graph. The two-way random intraclass correlation coefficient (absolute agreement), ICC (2, 2) was 0.82 with a 95% confidence interval of [0.80;0.83] indicating a good level of agreement.In a second step, these judgments were used to create numerical cutoffs to make the entire coding procedure automated and standardized. These cutoffs were then applied to the raw rating data and final data analysis was conducted on the coding determined by the numerical cutoffs only. The algorithms that were applied to the raw rating data to create the final coding assignments are outlined in Table 8.For example, if ratings decreased $>5$ points between the first and second rating (t2-t1 $< -5$) and did not increase $>5$ points between the second and third rating (t3-t2 $< 5$), then the pattern was coded as "decreasing", or a change away from a correct problem representation. If no change of $>5$ points was seen between any two ratings, then the pattern was given a "flat" coding. However, one exception was made for "flat" patterns that started at a very high initial level (within 5 points of the maximum rating of 100). They were coded as "high flat" patterns to separate them from all other flat patterns.If patterns increased by $>5$ points between any two ratings and never decreased $>5$ points, then the pattern was "increasing". These patterns were then further categorized as representing either an incremental or sudden large increase.[12] To receive a sudden coding, the pattern could only contain one large increase of 25 or greater, which could either occur between the first and second rating ("sudden increase before 2$^{nd}$ rating") or between the second and third rating ("sudden increase before 3$^{rd}$ rating"). If, however, there existed a second increase in target verb ratings of $>5$, an "incremental increase" coding was assigned. A second small

**Table 8**
Criteria for target verb rating pattern coding in Experiment 2.

| Pattern code | Starting point | Change between rating 1 and 2 | | Change between rating 2 and 3 |
|---|---|---|---|---|
| Decreasing | | t2-t1 $< -5$ | AND | t3-t2 $< 5$ |
| | | t2-t1 $< 5$ | AND | t3-t2 $< -5$ |
| Flat | (t1 $< 95$) AND | $\lvert$t2-t1$\rvert < 6$ | AND | $\lvert$t3-t2$\rvert < 6$ |
| Incremental increase | | $5 <$ t2-t1 $< 25$ | AND | t3-t2 $> -5$ |
| | | t2-t1 $> -5$ | AND | $5 <$ t3-t2 $< 25$ |
| Sudden increase before 3rd rating | | $\lvert$t2-t1$\rvert < 6$ | AND | t3-t2 $\geq 25$ |
| Sudden increase before 2nd rating | | t2-t1 $\geq 25$ | AND | $\lvert$t3-t2$\rvert < 6$ |
| High flat | (t1 $\geq 95$) AND | $\lvert$t2-t1$\rvert < 6$ | AND | $\lvert$t3-t2$\rvert < 6$ |

increase of 5 points or less was allowed for a sudden coding. All remaining patterns (consisting of a middle rating that was $>5$ points higher or lower than the first and last ratings, creating a zig zag pattern) were coded as "other".Table 9 shows the number of correct and incorrect solutions falling into each pattern category. A mixed effects model (binary logistic regression on correct solutions) with random intercepts for participants ($Z = 2.88, p = .004$), and tricks ($Z = 2.49, p = .013$) showed that the proportion of correct solutions did vary across pattern categories, $F(6, 1856) = 12.40$, $p < .001$. Pairwise comparisons revealed two different levels of solution success within the categories. As can be seen in Table 9, decreasing, flat and other patterns were more likely to be associated with incorrect solutions, and did not significantly differ from each other. In contrast, incremental increase, sudden increase, and high flat patterns were more likely to be associated with correct solutions, and did not significantly differ from each other. The significant difference in solution rates between flat and high flat patterns validates the decision to consider them separately.

*3.2.4.2. Does target verb rating pattern predict Aha! ratings?* In a prior study, Danek, Williams, & Wiley, 2020 found that tricks solved with a sudden increase in target verb ratings showed increased Aha! on correct solutions. Thus, the first analysis using target verb rating

[12] The cutoff value to discriminate between incremental and sudden increases in target verb ratings was based on judgments made by coders using visual analysis. They identified increasing patterns for which there was a single large jump between two ratings versus a more incremental increase across all three. There were 189 cases of increasing patterns that had one large jump of 25 or more. For 78.8% (149/189) of these cases, both raters agreed in categorizing the pattern as sudden. For cases that did not have one large jump of 25 or more, both raters agreed the pattern was not sudden on 91.2% (1527/1674) of cases, while a sudden pattern was perceived by both raters on 3.23% (54) of these cases. There were 0 cases with a jump of 25 or more that failed to be perceived as sudden by either rater. Therefore, the criterion for any increase between two consecutive target verb ratings to be coded as "sudden" was set to 25 or more. (A smaller jump would have reduced agreement to below 70%.)

**Table 9**
Pattern frequencies and number of subjects who contributed data to each category in Experiment 2.

| | Correct solutions | | Incorrect solutions | |
|---|---|---|---|---|
| | % of solutions assigned to this category (out of 933 correct solutions) | # of subjects contributing data to this category (out of 127) | % of solutions assigned to this category (out of 930 incorrect solutions) | # of subjects contributing data to this category (out of 127) |
| Decreasing patterns | 11.4% (106) | 68 | 20.5% (191) | 87 |
| Flat patterns | 12.1% (113) | 69 | 24.5% (228) | 95 |
| Incremental increase patterns | 22.8% (213) | 91 | 16.9% (157) | 82 |
| Sudden increase before 3rd rating | 2.7% (25) | 20 | 1.4% (13) | 12 |
| Sudden increase before 2nd rating | 10.6% (99) | 63 | 5.6% (52) | 44 |
| High flat patterns | 25.1% (234) | 79 | 11.4% (106) | 51 |
| Other | 15.3% (143) | 76 | 19.7% (183) | 87 |

*Note.* Actual observed frequencies out of 1863 observations (933 correct and 930 incorrect solutions) appear in parentheses.

pattern coding tested whether that result would replicate in the present data. A mixed effects model was calculated to test whether the 6 different patterns would be associated with differences in average Aha! ratings, with random intercepts for participants ($Z = 6.61$, $p < .001$), and tricks ($Z = 2.38$, $p = .017$), and fixed effects for pattern, solution correctness, and their dummy-coded interaction. As shown in Fig. 4, mean Aha! ratings varied as a function of both solution correctness ($F(1, 1525) = 63.34$, $p < .001$) and pattern ($F(5, 1525) = 2.48$, $p = .03$). The interaction was also significant, $F(5, 1525) = 3.55$, $p = .003$. To follow-up this interaction, the effects of pattern were examined separately for correct and incorrect solutions. No differences were seen in Aha! ratings due to pattern on tricks where incorrect solutions were generated, $F(5, 741) = 1.78$, $p = .115$. However, a significant effect of pattern was seen for correct solutions, $F(5, 784) = 3.42$, $p = .005$. Patterns showing one large jump between the first and second viewing led to significantly greater Aha! ratings than incremental increase, flat or decreasing patterns, replicating the prior finding of Danek, Williams, & Wiley, 2020.The remaining comparisons with the two new categories (high flat and sudden increase before third rating) showed that patterns with one large jump before the second rating led to greater Aha! experiences than high flat patterns, but not greater than patterns with a sudden increase before the third rating. Since for correct solutions, the sudden increase before third rating category consisted of only 25 observations (<3% of the data, see Table 9) coming from only 20 subjects, with high variance as seen in Fig. 4, this category remains difficult to interpret.

*3.2.4.3. Does target verb rating pattern predict recall of solution?* A novel question for Experiment 2 was whether differences in target verb rating patterns would be associated with differences in recall of solutions one week later. A mixed effects model (binary logistic regression on solution recall) was calculated to test whether the 6 different patterns would be associated with differences in memory for solutions, with random intercepts for participants ($Z = 4.36$, $p < .001$), and tricks ($Z = 2.12$, $p = .034$), and fixed effects for pattern, solution correctness, and their dummy-coded interaction. Number of observations was 1863. In this model, recall differed only by correctness of solution, $F(1, 1525) = 36.63$, $p < .001$, but not due to pattern, $F(5, 1525) = 1.26$, $p = .28$, nor was there a significant interaction, $F(5, 1525) = 0.21$, $p = .96$.As can be seen in Fig. 5, this lack of an effect due to pattern means there was no recall advantage for correct solutions that followed sudden increases toward a correct representation before the second rating (the leftmost pair of columns) compared to the three other patterns that were least likely to involve sudden restructuring (the three rightmost pairs of columns: incremental increase, flat and decreasing). If anything, high flat and sudden increases before the third rating showed the highest likelihood of solution recall. Given that both these categories likely include different types of solutions, including some sudden, these two patterns are not collapsed with either the sudden increase before the second rating pattern or the remaining non-sudden patterns for later analyses. A *sudden restructuring* measure was derived from a contrast code using 1 for "sudden" (containing only patterns with sudden increases before the second rating) and 0 for "not sudden" (incremental increases, flat, decreasing patterns).

*3.2.4.4. Does change toward a correct problem representation predict Aha! ratings and recall of solution?* As a second measure of representational change, the magnitude of overall *change toward a correct problem representation* was computed from differences between the initial and final importance-to-solution ratings using both the target

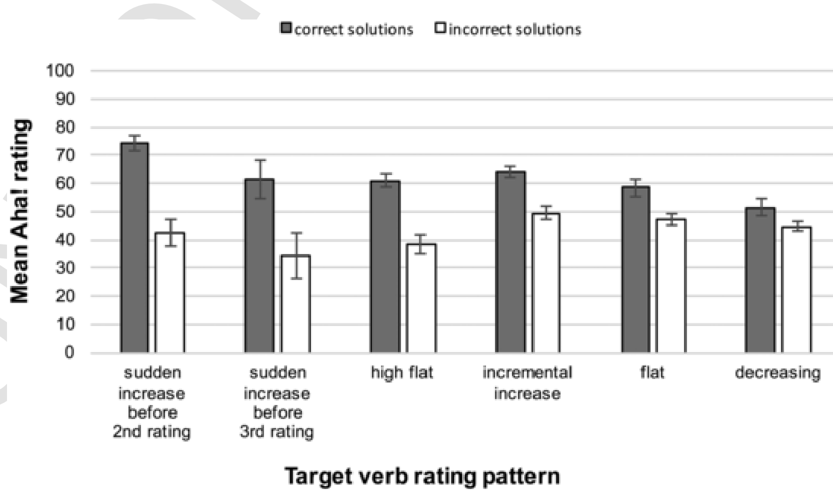


**Fig. 4.** Mean Aha! ratings for correct and incorrect solutions as a function of target verb rating pattern in Experiment 2. Error bars denote standard error of the mean.

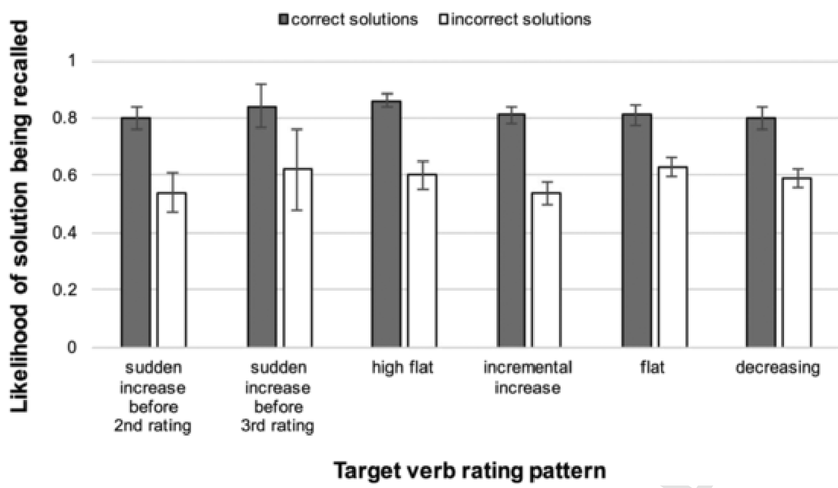


**Fig. 5.** Mean recall rates for correct and incorrect solutions as a function of target verb rating pattern in Experiment 2. Error bars denote standard error of the mean.

verbs and the distractor verbs. In contrast to the previous analyses, this second measure does not contain any information about the "shape" of the change in target verb ratings over time. Following Cushen and Wiley (2012), a difference score reflecting the correctness of the problem representation was computed at both the initial and final rating timepoints. The difference score was created by subtracting the average ratings of the distractor verbs from the rating of the target verb for each trick. This provides a score ranging from −1 (completely incorrect representation) to 1 (correct representation), where 0 implies no distinction in importance ratings between the target and distractor verbs. Then the magnitude of overall *change toward a correct problem representation* was computed by subtracting the difference scores computed at the initial timepoint from those at the final timepoint. Larger values of this measure reflect greater overall change toward a correct problem representation.The third measure of representational change, the magnitude of overall *change away from an incorrect problem representation,* was computed as a complement to this measure. A parallel set of analyses were done, this time using importance-to-solution ratings on the biased verb and the distractor verbs. The relation of these two additional measures of representational change to the *sudden restructuring* contrast, Aha! ratings, correct solution, and recall are shown in Table 10. Both *sudden restructuring* and greater overall *change toward a correct problem representation* were significantly associated with stronger Aha! ratings, while greater overall *change away from an incorrect problem representation* was not. Similarly, only *sudden restructuring* and greater overall *change toward a correct problem representation* were significantly associated with generating a correct solution, as well as with feelings of certainty. Finally, only one of the three measures, magnitude of overall *change toward a correct problem representation,* was correlated with solution recall. The *sudden restructuring* contrast did not correlate with solution recall, nor did it correlate with solvers' self-reported perceptions of suddenness (or surprise) in their own solution progress.

**Table 10**
Simple correlations between sudden restructuring, change toward a correct problem representation, change away from an incorrect problem representation, and other measures in Experiment 2.

| | Sudden restructuring | Change toward correct representation | Change away from incorrect representation |
|---|---|---|---|
| Aha! | 0.12** | 0.18** | 0.05 |
| Pleasure | 0.05 | 0.09** | 0.02 |
| Surprise | −0.01 | 0.02 | 0.01 |
| Suddenness | 0.00 | −0.02 | 0.00 |
| Relief | 0.06* | 0.05 | 0.06* |
| Certainty | 0.12** | 0.16** | 0.02 |
| Drive | 0.02 | 0.04 | −0.01 |
| Correctness | 0.15** | 0.23** | 0.03 |
| Recall | 0.02 | 0.07* | 0.00 |
| Sudden restructuring | – | 0.57** | 0.02 |
| Change toward correct representation | – | – | −0.08** |

*Note.* $N = 1159$.
* $p < .05$.
** $p < .01$.

### 3.2.5. *Do affective or cognitive components predict the insight memory advantage?*

The preceding analyses identified one measure of representational change (overall *change toward a correct representation*) and two specific underlying dimensions of the Aha! experience (pleasure and certainty) that uniquely predicted recall. The simple correlations among these predictors and other measures are shown in Table 10.

The goal for the final analysis was to combine both affective and cognitive predictors from the prior analyses within the same models to test which might uniquely drive the *insight memory advantage*. A mixed effects model (number of observations = 1159) was used to perform a binary logistic regression on recall to test whether affective or cognitive components would be associated with differences in memory for solutions, with random intercepts for participants ($Z = 4.01$, $p < .001$), and tricks ($Z = 1.86$, $p = .063$), and fixed effects for solution correctness, overall *change toward a correct problem representation*, pleasure and certainty ratings. The overall model was significant, $F(4, 1154) = 23.95$, $p < .001$. As shown in Table 11, solution correctness, pleasure and certainty all were significant unique predictors of recall, while overall *change toward a correct representation* did not account for any unique variance once solution correctness was accounted for.

Although the *sudden restructuring* measure did not show a simple correlation with recall, it was theoretically interesting to directly test whether it would predict the *insight memory advantage* as part of the full model. Thus, the same analysis was re-run, only using *sudden restructuring* as a predictor instead of the magnitude of overall change toward a correct representation. As shown in Table 12, the results were largely the same, as correctness, certainty, and pleasure (at $p = .05$) continued to predict recall, but *sudden restructuring* did not.

**Table 11**
Binary logistic regression model (mixed effects) with solution correctness, change toward correct representation, pleasure and certainty as predictors of solution recall.

| | Unstandardized coefficient B | SE B | t | p | d |
|---|---|---|---|---|---|
| Constant | 0.553 | 1.270 | | | |
| Correctness | 0.807 | 0.172 | 4.70 | $p < .001$ ** | 0.28 |
| Change toward correct representation | 0.000 | 0.002 | 0.15 | $p = .88$ | 0.01 |
| Pleasure | 0.007 | 0.004 | 1.98 | $p = .047$ * | 0.12 |
| Certainty | 0.017 | 0.003 | 5.61 | $p < .001$ ** | 0.33 |

*Note.* N = 1159.
* $p < .05$.
** $p < .01$.

**Table 12**
Binary logistic regression model (mixed effects) with solution correctness, sudden restructuring, pleasure and certainty as predictors of solution recall.

| | Unstandardized coefficient B | SE B | t | p | d |
|---|---|---|---|---|---|
| Constant | 0.567 | 1.270 | | | |
| Correctness | 0.822 | 0.171 | 4.81 | $p < .001$ ** | 0.28 |
| Sudden restructuring | −0.131 | 0.229 | −0.57 | $p = .567$ | 0.03 |
| Pleasure | 0.007 | 0.004 | 1.97 | $p = .050$ | 0.12 |
| Certainty | 0.017 | 0.003 | 5.65 | $p < .001$ ** | 0.33 |

*Note.* N = 1159.
* $p < .05$.
** $p < .01$.

### *3.3. Discussion*

These results indicate that feelings of certainty and pleasure along with reaching a correct solution are the key dimensions of the problem solving experience that independently benefit the *insight memory advantage*. The contribution of pleasure to the *insight memory advantage*, regardless of correctness of solution, provides clear evidence for an affective component. At the same time, the observed benefit from having generated a correct solution, over and above the effect of pleasure, provides some evidence for a cognitive component. The unique contributions from pleasure and correct solution suggest that both affective and cognitive components are responsible for the *insight memory advantage*. However, no evidence was found for either the amount of change toward a correct problem representation nor for suddenness in the change toward a correct representation mattering after correctness of solution was accounted for.

## 4. General discussion

The *insight memory advantage* is the finding that solutions that are generated with an Aha! experience are better remembered than those which lack this feeling. This effect was replicated in both of the present studies. Effect sizes place this memory phenomenon as comparable to effects such as survival processing, while not as strong as other effects such as retrieval practice. Beyond simply replicating the basic effect established in prior work, the main goal of this research was to understand the extent to which affective and cognitive components of solutions contribute to the *insight memory advantage*. In particular, it tested whether the memory advantage is modulated by the affective component (the strong feelings that typically accompany the Aha! experience), or by the cognitive component (the restructuring or change within the problem representation that occurs during insightful problem solving), or whether both are required for the memory advantage to occur (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013; Gick & Lockhart, 1995; Kizilirmak et al., 2016). The first study showed the connection of the *insight memory advantage* to correctly generated solutions, subjective Aha! experiences, and feelings of certainty as seen in Danek, Fraps, von Müller, Grothe, & Öllinger, 2013. The second study identified feelings of pleasure as an additional unique predictor of solution recall.

The results of both studies provided evidence to support the role of the cognitive component, as memory for solutions was in part determined by whether solvers had reached a correct solution. Yet, if the memory advantage were truly an "insight" memory advantage, where a "true insight" is marked by a sudden shift in representation, then target verb rating patterns suggesting sudden restructuring should have led to higher recall rates for correct solutions. Instead, sudden restructuring was not found to relate to recall. In fact, sudden and incremental increase patterns, but also flat and decreasing patterns led to nearly identical recall rates. Thus, the more specific hypothesis that sudden representational change might play a role in the *insight memory advantage* was not supported.

The results also provided evidence to support the role of the affective component in the *insight memory advantage*. The results of the current studies are consistent with the speculation based on prior results (Danek et al., 2014a; Danek & Wiley, 2017; Kizilirmak et al., 2016) that the unique variance explained by the Aha! ratings for solution memory over and above the variance explained by reaching a correct solution might be best represented as a benefit of pleasure or positive affect on memory. In particular, the results of the second experiment, which showed a unique contribution of pleasure ratings over and above correct solution and feelings of certainty, provide support for the role of positive affect in supporting better solution memory.

Several physiological mechanisms, including amygdala activation, have been proposed as potential explanations for the effect that emotion seems to have on memory (Hamann et al., 1999). There are now several studies that have linked insight problem solving to amygdala activity. Not only is it correlated with self-reports of insight when people solve compound remote associates problems (Jung-Beeman et al., 2004), but amygdala activity has also been linked to memory for solutions in a study where insight was induced by presenting solutions to those problems (Kizilirmak et al., 2019) as well as in a perceptual induced-insight paradigm (Ludmer et al., 2011). It seems plausible that experiencing an Aha! while solving a problem (or even while being shown the solution) constitutes an emotional event that activates the amygdala which in turn can lead to better memory for the solution. Yet, further work is needed to test these speculations.

With regard to generalizability, the *insight memory advantage* has so far appeared in two rather different task domains: Discovering the secret method behind magic tricks (in Danek, Fraps, von Müller, Grothe, & Öllinger, 2013, and in the present studies), and identifying objects in degraded images (Kizilirmak et al., 2016). Solving these problems elicits strong pleasurable reactions, and in particular the magic tricks seem quite close to "real world" problem solving as it happens outside of the psychologist's laboratory. This sets the ground for future work possibly extending the memory effect to other task domains.

There were a number of other ancillary findings that demonstrate the relation of the present studies to prior work, and help to contribute to the research base. One set of ancillary findings concerns the Aha! ratings themselves. Across two experiments, higher Aha! ratings led to better recall for solutions after one week. This is in accordance with earlier findings (Danek, Fraps, von Müller, Grothe,

& Öllinger, 2013; Kizilirmak et al., 2016). However, in both experiments, the memory advantage was not confined to correct solutions. The lack of an interaction between Aha! ratings and correctness is important because it suggests that regardless of the correctness of the solution, the feelings underlying the Aha! experience may be in part responsible for the stronger memory traces. The results thus extend prior work showing that incorrect solutions can sometimes prompt strong Aha! experiences (false insights, see Danek & Wiley, 2017), and that when they do, that also improves their memorability even though they do not represent correct or insightful solutions.

Another set of findings highlights the special status afforded to correct solutions. Across both experiments, correct solutions were better remembered than incorrect solutions, even in the absence of feedback. One possible explanation for this effect could be differences between correct and incorrect solutions with regard to content. The majority of the magic tricks used in the present experiments (14 of 18) had a solution which consisted of only one step (i.e., the correct solution could be explained by one piece of information or one hand movement of the magician). In contrast, the incorrect solutions suggested by participants were often less elegant, consisting of multiple different pieces of information, with one or more of them being implausible. Correct solutions are more complete and holistic, making everything fall into place, and they have a good *Gestalt* in the sense of Wertheimer (1925), as also argued by Danek & Salvi, 2020. These qualitative differences could explain why correct solutions were remembered better. They may also explain the finding of higher certainty ratings for correct vs. incorrect solutions (Ackerman & Zalmanov, 2012; Danek et al., 2014b; Hedne et al., 2016; Threadgold et al., 2018) as well as the repeated finding across task domains that problem solvers give higher Aha! ratings for correct solutions than for incorrect ones (Danek et al., 2014b; Danek, Williams, & Wiley, 2020; Danek & Wiley, 2017; Hedne et al., 2016; Salvi et al., 2016; Threadgold et al., 2018; Webb et al., 2016, 2017; Webb et al., 2019a). Whether this is best considered as an *accuracy effect* or a *correctness effect* may depend upon the dependent variable (Threadgold et al., 2018). For a review on existing data and a discussion of this effect in insight problem solving, see and for a different view and a discussion that reaches outside of the problem solving domain, see Webb et al. (Webb et al., 2019b).

An "intuitive sense of success" (Gick & Lockhart, 1995, p. 215) was also observed in the present studies. Solvers' subjective feelings of certainty not only predicted the actual correctness of their solutions, but also predicted Aha! experiences as well as solution memory. Interestingly, strong feelings of certainty were not specific to correct solutions, and sometimes participants felt certain that they had reached a correct solution even when they had not which constitutes a high confidence error (Butterfield & Metcalfe, 2001). The lack of an interaction between certainty ratings and solution correctness in predicting solution recall suggests that regardless of correctness, feelings of certainty or confidence were in part responsible for the stronger memory traces, even when they were illusory.

A final set of findings are relevant for the question of how to best assess the cognitive component of sudden restructuring during problem solving. It is important to emphasize again that the present approach relied on situational judgments related to solvers' problem representations, and not metacognitive judgments about their progress (as distinguished by Ash and Wiley, 2008). Similar to the results of Cushen and Wiley (2012), the sudden restructuring measure did not correlate with subjective, metacognitive ratings of suddenness (and surprise). There are a number of potential reasons why metacognitive ratings may not be valid measures of actual solution progress. The validity of "suddenness" ratings may suffer from confusion about what "suddenness" means. (Many students seem to endorse this rating when they solve a problem quickly, as opposed to when the solution pops into mind all at once). Of course, there is the potential for other feelings to leak into the "suddenness" rating. And, in the present study, the "suddenness" ratings could also have been affected by solvers being forced to view each trick 3 times. Alternatively, solvers may simply lack accurate metacognitive awareness of their solution processes (Cushen & Wiley, 2012; Ellis et al., 2011; Metcalfe, 1986). However, a recent study using the matchstick arithmetic task found a close correspondence between a more objective and fine-grained measure of solution progress (eye-tracking) and subjective ratings of suddenness and surprise (Bilalić, Graf, Vaci, & Danek, 2019). In that study, solvers who began to look earlier at the crucial solution element rated their solution as less sudden and less surprising than solvers who attended to it only shortly before solution.

On the other hand, although the sudden restructuring measure derived in this study did not predict memory for solutions, it did predict Aha! experiences, replicating Danek, Williams, & Wiley, 2020. Sudden increases in target verb ratings between the first and second viewing of a trick led to higher Aha! ratings than all other non-sudden patterns (for correct solutions). This finding provides further evidence for the theoretically assumed relationship between affective (Aha!) and cognitive (sudden restructuring) aspects of insightful problem solving, and shows the utility of using importance-to-solution ratings to assess problem representations. It is in accordance with another recent study (Kizilirmak et al., 2018) where feeling-of-warmth ratings for problems solved with Aha! showed more of a sudden increase shortly before solution, as compared to problems solved without Aha! which showed more of a gradual increase. At the same time, both of these measures of restructuring (sudden increases in verb ratings and feeling-of-warmth ratings) are necessarily indirect ones. It remains a challenging question how researchers might best assess an individual's mental representation of a problem and detect changes in it (as opposed to affective reactions that may be manifest in physiological trace measures such as pupil dilation, heart rate, or grip strength). Future work is needed using other methods to provide converging sources of evidence for discontinuity during the solution process, which might allow researchers to be sure they are in a position to test for effects associated with sudden restructuring. It is possible that the failure to find an effect of sudden restructuring on solution memory in this study was because we did not have a sensitive enough measure of discontinuities or abrupt shifts in representation. Alternatively, the failure to find such an effect brings into question whether suddenness should be viewed as an essential feature of the insight experience.

## 5. Conclusions

In sum, the explanation for the memory advantage that seems best supported by this data is that it is the joint consequence of finding correct solutions, a subjective feeling that one has found a correct solution, and experiencing an emotional pleasurable reaction during the problem solving process, that all contribute to better memory for the solution. Thus, the present findings suggest that the previously coined *insight memory advantage* is in part an affective advantage and in part a cognitive advantage. It does seem to depend on having actually solved a problem correctly, but does not seem to rely on having reached the solution via a sudden restructuring process.

## Acknowledgements

This work was supported by the DFG (German Research Foundation) with a grant to Amory Danek [grant number DA 1683/1-1]. We are indebted to magician Thomas Fraps (http://www.thomasfraps.com) for performing the magic tricks used in this study. For help with data collection and coding, we thank Shannon Menard and Jocelyn Rodriguez (Experiment 1) and Karen Reyna and Oscar Tellez (Experiment 2). We also thank Franziska Konitzer who wrote the Psychopy code for both experiments. We are grateful to Simon J. Cropper and two anonymous reviewers for their helpful comments on an earlier version of this article.

## Supplementary material

The datasets of the present experiments are available under doi: http://dx.doi.org/10.23668/psycharchives.3115 at the open repository PsychArchives https://www.psycharchives.org.

## References

Abbott, A (2005). Science as illusion. Nature, 434, 820.
Ackerman, R, & Zalmanov, H (2012). The persistence of the fluency–confidence association in problem solving. Psychonomic Bulletin & Review, 19(6), 1187–1192.
Ash, I K, Cushen, P J, & Wiley, J (2009). Obstacles in investigating the role of restructuring in insightful problem solving. The Journal of Problem Solving, 2(2), 6–41.
Ash, I K, Jee, B, & Wiley, J (2012). Investigating insight as sudden learning. The Journal of Problem Solving, 4(2), 150–176.
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.
Auble, P M, Franks, J J, & Soraci, S A (1979). Effort toward comprehension: Elaboration or "aha!"? Memory & Cognition, 7(6), 426–434.
Baayen, R H, Davidson, D J, & Bates, D M (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390–412.
Bilalić, M, Graf, M, Vaci, N, & Danek, A H (2019). The temporal dynamics of insight problem solving – restructuring might not always be sudden. Thinking & Reasoning.
Butterfield, B, & Metcalfe, J (2001). Errors committed with high confidence are hypercorrected. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(6), 1491–1494.
Cosmelli, D, & Preiss, D D (2014). On the temporality of creative insight: A psychological and phenomenological perspective. Frontiers in Psychology, 5, 1184.
Cranford, E A, & Moss, J (2012). Is insight always the same? A protocol analysis of insight in compound remote associate problems. The Journal of Problem Solving, 4(2), 128–153.
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.
Danek, A H (2018). Magic tricks, sudden restructuring and the Aha! experience: A new model of non-monotonic problem solving. In Vallée-Tourangeau, F (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 (2013). Aha! experiences leave a mark: Facilitated recall of insight solutions. Psychological Research, 77(5), 659–669.
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.
Danek, A H, Fraps, T, von Müller, A, Grothe, B, & Öllinger, M (2014). Working wonders? Investigating insight with magic tricks. Cognition, 130(2), 174–185.
Danek, A H, & Salvi, C (2020). Moment of truth: Why Aha! experiences are correct. The Journal of Creative Behavior, 54, 484–486. 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.
Danek, A H, Williams, J, & Wiley, J (2020). Closing the gap: Connecting sudden representational change to the subjective Aha! experience in insightful problem solving. Psychological Research, 84, 111–119.
Dominowski, R L, & Buyer, L S (2000). Retention of problem solutions: The re-solution effect. The American Journal of Psychology, 113(2), 249–274.
Dominowski, R L, & Dallob, P (1995). Insight and problem solving. In Sternberg, R J, & Davidson, J E (Eds.), The nature of insight (pp. 33–62). Cambridge, MA: MIT Press.
Duncker, K (1945). On problem-solving. Psychological Monographs, 58. (Whole No. 270).
Durso, F T, Rea, C B, & Dayton, T (1994). Graph-theoretic confirmation of restructuring during insight. Psychological Science, 5(2), 94–97.
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.
Gick, M L, & Lockhart, R S (1995). Cognitive and affective components of insight. In Sternberg, R J, & Davidson, J E (Eds.), The nature of insight (pp. 197–228). Cambridge, MA: MIT Press.
Gruber, H E (1995). Insight and affect in the history of science. In Sternberg, R J, & Davidson, J E (Eds.), The nature of insight (pp. 397–431). Cambridge, MA: MIT Press.
Grunewald, K, & Bowden, E M (2018). Whose insight is it anyway? In Vallée-Tourangeau, F (Ed.), Insight: On the origins of new ideas (pp. 28–50). London: Routledge.
Hamann, S B, Ely, T D, Grafton, S T, & Kilts, C D (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. Nature Neuroscience, 2(3), 289–293.
Hebb, D O (1949). The organization of behavior. New York: John Wiley & Sons.
Hedne, M R, Norman, E, & Metcalfe, J (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. Frontiers in Psychology, 7, 1314.
Jung-Beeman, M, Bowden, E M, Haberman, J, Frymiare, J L, Arambel-Liu, S, Greenblatt, R, & Reber, P J, et al. (2004). Neural activity when people solve verbal problems with insight. PLoS Biology, 2(4), 500–510.
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.
Kizilirmak, J M, Schott, B H, Thuerich, H, Sweeney-Reed, C M, Richter, A, Folta-Schoofs, K, & Richardson-Klavehn, A (2019). Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. Consciousness and Cognition, 69, 113–132.
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.
Köhler, W (1921). Intelligenzprüfungen am Menschenaffen. Berlin: Springer.
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.
Ludmer, R, Dudai, Y, & Rubin, N (2011). Uncovering camouflage: Amygdala activation predicts long-term memory of induced perceptual insight. Neuron, 69(5), 1002–1014.
Maier, N R F (1945). Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. Journal of Experimental Psychology, 35(5), 349.
Metcalfe, J (1986). Premonitions of insight predict impending error. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(4), 623–634.
Metcalfe, J, & Wiebe, D (1987). Intuition in insight and noninsight problem solving. Memory & Cognition, 15, 238–246.
Ohlsson, S (1992). Information-processing explanations of insight and related phenomena. In Keane, M T, & Gilhooly, K J (Eds.), Advances in the psychology of thinking (Vol. 1, pp. 1–44). London: Harvester-Wheatsheaf.
Osgood, C E (1953). Method and theory in experimental psychology. New York: Oxford University Press.
E. Page-Gould Multilevel modeling workshopRetrieved fromhttp://page-gould.com/mlm/mpa/2013
Rosenthal, R, & Rubin, D B (2003). r equivalent: A simple effect size indicator. Psychological Methods, 8(4), 492–496.
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.
Scheerer, M (1963). Problem-solving. Scientific American, 208, 118–128.
Schooler, J W, & Melcher, J (1995). The ineffability of insight. In Smith, S M, Ward, T B, & Finke, R A (Eds.), The creative cognition approach (pp. 97–133). Cambridge, MA: MIT Press.
Schwarz, N, & Clore, G L (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology, 45(3), 513–523.
Schwarz, N, & Clore, G L (2003). Mood as information: 20 years later. Psychological Inquiry, 14(3–4), 296–303.
Thagard, P, & Stewart, T C (2011). The Aha! Experience: Creativity through emergent binding in neural networks. Cognitive Science, 35(1), 1–33.
Threadgold, E, Marsh, J E, & Ball, L J (2018). Normative data for 84 UK English rebus puzzles. Frontiers in Psychology, 9.
Topolinski, S, & Reber, R (2010). Gaining insight into the "aha" experience. Current Directions in Psychological Science, 19(6), 402–405.
Webb, M E, Cropper, S J, & Little, D R (2019). "Aha!" is stronger when preceded by a "huh?": Presentation of a solution affects ratings of aha experience conditional on accuracy. Thinking & Reasoning, 25(3), 324–364.
Webb, M E, Laukkonen, R E, Cropper, S J, & Little, D R (2019). Commentary: Moment of (perceived) truth: Exploring accuracy of Aha!. The Journal of Creative Behavior: Experiences.
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.
Webb, M E, Little, D R, & Cropper, S J (2017). Once more with feeling: Normative data for the aha experience in insight and noninsight problems. Behavior Research Methods, 1–22.
Wertheimer, M (1925). Über Schlussprozesse im produktiven Denken. In Wertheimer, M (Ed.), Drei Abhandlungen zur Gestalttheorie (pp. 164–184). Erlangen: Verlag der Philosophischen Akademie.
Wertheimer, M (1959). Productive thinking. New York: Harper.
Woodworth, R S, & Schlosberg, H (1954). Experimental psychology. New York: Holt.