Thinking & Reasoning

ISSN: 1354-6783 (Print) 1464-0708 (Online) Journal homepage: https://www.tandfonline.com/loi/ptar20

Problem solving of magic tricks: guiding to and through an impasse with solution cues

Judit Pétervári & Amory H. Danek

To cite this article: Judit Pétervári & Amory H. Danek (2019): Problem solving of magic tricks: guiding to and through an impasse with solution cues, Thinking & Reasoning, DOI: 10.1080/13546783.2019.1668479

To link to this article: https://doi.org/10.1080/13546783.2019.1668479

Published online: 03 Oct 2019.

THINKING & REASONING
https://doi.org/10.1080/13546783.2019.1668479

ARTICLE

Problem solving of magic tricks: guiding to and through an impasse with solution cues

Judit Pétervári[a] and Amory H. Danek[b]*

[a]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK; [b]Division of Neurobiology, Department Biology II, Ludwig-Maximilians-Universität München, Munich, Germany

ABSTRACT
This study investigated how problem solvers get into and out of a state of impasse while solving difficult problems. 47 participants had to decipher the secret method behind 33 magic tricks while repeatedly giving feeling of warmth ratings. After the first viewing of each trick, participants were led into an impasse by presenting two implausible solutions, together with the information that those were incorrect. After another viewing, cues were given to guide out of the impasse. Warmth ratings were flat and non-increasing after the implausible solution manipulation, suggesting a state of impasse. Cues helped to overcome the impasse, with higher solution rates for pictorial (49%) than for verbal cues (39%) and both higher than a no cue condition (29%). Warmth ratings also reflected cue efficacy, with higher ratings after helpful cues. This study represents a first attempt at influencing the onset and offset of the state of impasse.

ARTICLE HISTORY Received 1 February 2018; Accepted 6 September 2019

KEYWORDS Problem solving; insight; impasse; magic; feeling of warmth; cues

Introduction

Although the human brain continuously strives to make accurate predictions (Bar, 2007), many important discoveries occur unpredictably. Often problem solvers cannot reliably foresee whether or when they will be able to solve a difficult problem. The prepared mind perspective suggests that if they already possess all necessary information for reaching a solution, this can happen spontaneously (Seifert, Meyer, Davidson, Patalano, & Yaniv, 1995). Instead of relying on trial-and-error search, problem solvers may find the solution suddenly and all at once, and feel certain about being correct. This phenomenological experience is called an insight (e.g. Duncker, 1935).

CONTACT Judit Pétervári j.petervari@qmul.ac.uk
*Present address: Experimental and Theoretical Psychology, University of Heidelberg, Heidelberg, Germany

During the preceding, sometimes lengthy solving process, a period of impasse, defined as a state of feeling stuck with no visible steps towards the solution, is often encountered (Ohlsson, 1992).

The aim of this study is to investigate how problem solvers get into and then out of a state of impasse while trying to solve difficult problems. First, we increased the likelihood of reaching an impasse during the problem-solving process by taking away trivial, implausible solution ideas. Second, we examined which types of cues were the most efficient in guiding problem solvers out of the impasse, towards the solution of the problem.

On the behavioural level, being in a state of impasse is characterised either as inactivity or as perseverating behaviour, repeating previous moves and trying the same failing strategies over and over again (Fedor, Szathmáry, & Öllinger, 2015). On the phenomenological level, solvers experience a feeling of being stuck and not coming closer to solution. The present study focussed on this metacognitive feeling of not progressing towards a solution, and therefore operationalised impasse as non-increasing feeling of warmth ratings.

Theoretically, experiencing an impasse has often been hypothesised to be necessary for insight to occur (Ohlsson, 1992; Öllinger, Jones, & Knoblich, 2014b; Ormerod, MacGregor, & Chronicle, 2002; Seifert et al., 1995). The Representational Change Theory (RCT; Ohlsson, 1984, 1992, 2011) explains both the onset and the offset of an impasse (but not the precise timing of these events). At first, solving attempts are unsuccessful because of an incorrect representation of the problem, which leads the problem solver astray. If the evoked representation does not comprise the critical solution information and problem solvers are blocked by misleading assumptions or constraints imposed by their prior knowledge, they keep exploring the problem space but eventually stumble and face the state of impasse. The impasse can only be overcome if the biased problem representation is fundamentally changed ("representational change" in terms of Ohlsson, 1992). Ohlsson also postulated three hypothetical processes for how a representational change could be achieved, namely elaboration (extending the initial problem representation by e.g. detecting formerly overlooked details), re-encoding (e.g. a figure-ground reversal in geometry problems) and constraint relaxation (overturning implicitly imposed constraints, see Knoblich, Ohlsson, Haider, & Rhenius, 1999, for a more detailed account). The present work focuses on the latter process to break the impasse, implementing cues specifically designed to target constraints and to lead out of the impasse by helping problem solvers relax those constraints.

Empirically, what is known about the state of impasse? Beeftink, van Eerde, and Rutte (2008) conducted one of the few studies which obtained self-reports about the occurrence of impasse during the solving process and report that, on average, almost each participant experienced an impasse while solving cryptic crossword puzzles. Fedor et al. (2015) had participants press a button whenever they felt stuck on Katona's Five-Square problem (Katona, 1940) and found that 33% of solvers and 79% of non-solvers indicated impasse at least once. Further, solvers followed the classic impasse-insight sequence in about 50% of the cases, while the remaining participants had more complex sequences, with iterative stages of search and impasse, as theoretically postulated by Öllinger et al. (2014b). From video recordings of participants' behaviour while solving a variant of the 8 Coin Problem (Ormerod et al., 2002), Sas, Luchian, and Ball (2010) report that solvers experienced longer time periods of impasse, but also more instances of impasse than non-solvers (note that this is at odds with Fedor et al., 2015, who found less impasse for solvers).

The outlined dynamics of the hypothetical impasse-insight sequence are also supported by eye movement data, for example, Knoblich et al. (Knoblich, Ohlsson, & Raney, 2001) found that there was a period of impasse (defined as ostensible inactivity with few or no eye movements) which was then followed by solvers undergoing a shift in attention towards the crucial element of a matchstick problem and non-solvers remaining fixated on irrelevant problem elements.

In think aloud protocols, Schooler and Melcher found more statements indicating impasse such as "I just can't imagine … " or "I am just wondering where to go from here" (Schooler & Melcher, 1995, p. 115) when people solved insight problems as compared to when they solved analytical problems. Contrary to this, Cranford and Moss (2012) found similar rates of impasse for solutions classified as insight vs. solutions classified as non-insights, in think aloud protocols of compound remote associate (CRA) problems (Bowden & Jung-Beeman, 2003). They concluded that CRA problems may be less likely to induce impasse than other tasks, because they offer more options for generating new solution candidates (a new word) so that lengthy periods of impasse are less likely. As long as participants can harness the power of trial-and-error, and thus seemingly make progress, they will not assume to be in an impasse. The present study tried to counter-act this by actively eliminating often-mentioned false solution candidates, as described later.

However, despite the reviewed evidence and although impasse is often postulated to be a pre-condition for insight to occur (Ohlsson, 1992; Öllinger et al., 2014b; Ormerod et al., 2002; Seifert et al., 1995), there are also strong indications that solvers can experience insights without undergoing a previous impasse. If we look at the numbers reported previously, clearly impasse does not occur before every solution. For example, in Fedor et al.'s study (2015), impasse was reported by only 33% of solvers. Fleck and Weisberg identified instances of impasse in only 50% of think-aloud protocols (Fleck and Weisberg, 2013). In another think-aloud study, Cranford and Moss (2012) found that immediate solutions which are found right away, a few seconds after problem presentation, were nevertheless labelled as insight solutions by solvers, probably simply because the answer was sudden. However, there was no evidence of any other characteristic of insight such as impasse or restructuring (note that in these few seconds hardly any protocol data could be produced). Phenomenologically, when problem solvers were retrospectively asked about their Aha! experiences, the dimension of impasse was less strongly endorsed than the aspects of pleasure, suddenness or certainty (Danek, Fraps, von Müller, Grothe, & Öllinger, 2014a).

We conclude from the reviewed literature that there is evidence that impasse exists, that it can occur as part of the insight problem solving process, but that it is not a necessary condition for insight to take place (as already proposed by Danek, 2018). This conclusion has also been drawn by others (Derbentseva, 2007; Kounios & Beeman, 2014). In this light, it seems important to learn more about the conditions which may trigger a state of impasse and also ways to resolve the impasse once it has occurred. The present study thus represents an attempt at manipulating the onset and offset of impasse.

FoW ratings as a measure of impasse

To track the individual problem-solving experience, i.e. where participants are in the impasse-insight sequence, we followed Metcalfe and Wiebe's (1987) pioneering study protocol, who set up a differentiation between incremental and discontinuous problem-solving processes with the help of feeling of warmth (FoW) ratings. These ratings are subjective reports of how close participants feel to the solution and were invented based on a game in which a hidden object must be found by relying on heat-related feedback about how close one is standing to it. To yield a continuous measure of how people solve problems, they are typically taken repeatedly, at several time points during the solution process. FoW ratings have been used to investigate insight across many different task domains: Anagrams (e.g. Metcalfe, 1986), magic tricks (Hedne, Norman, & Metcalfe, 2016), CRA problems (Kizilirmak et al., 2018), 9 Dot Problem (e.g. Chein, Weisberg, Streeter, & Kwok, 2010), verbal puzzles (e.g. Laukkonen & Tangen, 2018; Metcalfe & Wiebe, 1987), spatial insight problems (e.g. Metcalfe & Wiebe, 1987), Triangle of Coins (e.g. Jausovec & Bakracevic, 1995; Metcalfe, 1986).

It is important to delineate what is (and what is not) measured by FoW ratings. As a sort of trace data, warmth ratings offer the possibility to track the dynamics of participants' solution process online, i.e. while they work on the problem. They are metacognitive judgements (as opposed to situational judgements, see Ash & Wiley, 2008, for a distinction) and measure subjective, not objective progress on the problem. In the present study, we used FoW ratings to measure when participants feel stuck on a problem. Being in a state of impasse was operationalised as flat, non-increasing FoW ratings. If the impasse is broken, FoW ratings should increase. However, warmth ratings do not measure any affective dimensions of insight such as pleasure, relief or drive.

Recently, Laukkonen and Tangen (2018) advocated for using a post-hoc self-report measure of the Aha! experience ("Did you have an Aha! moment?"), as proposed by Bowden, Jung-Beeman, Fleck, and Kounios (2005), instead of FoW ratings. While we fully agree with them that post-hoc Aha! judgements are very valuable measures and have also advocated for their use (e.g. Danek, 2018; Danek et al., 2014a), we think it would be a mistake to dismiss FoW ratings as a measure of insight. The reasoning in the Laukkonen study was that, if the two measures are concurrently assessed, but do not converge, they must be measuring different aspects of insight. The Aha! experience is typically conceptualised as the affective component of insight (e.g. Danek, Williams, & Wiley, 2018; Gick & Lockhart, 1995), as a striking, positive emotional response to finding the solution. There is evidence that the self-reported Aha! experience is a multi-dimensional construct with pleasure, suddenness in the emergence of a solution and certainty or confidence about the found solution as key components (Danek & Wiley, 2017). This is different from FoW ratings. Thus, from a conceptual standpoint, although both measures touch on the issue of the suddenness of insight, it seems clear that warmth ratings and post-hoc Aha! self-reports were never meant to capture the same construct. Laukkonen and Tangen (2018) found that if the patterns of FoW ratings indicated a sudden solution, then there was a 75% chance that the self-reports also indicated an insight (operationalized as an Aha! experience which in turn was described as a feeling of suddenness in the emergence of solution, together with a feeling of obviousness, i.e. certainty about solution). This speaks for a connection between warmth ratings and post-hoc self-reports of insight. On the other hand, if warmth ratings did not indicate a sudden solution, there was still a chance of 50% that participants would self-report an insight. This speaks against such a connection. In addition to these unclear findings, some methodological issues may limit the conclusions that can be inferred from this study: First, only 30% of trials were used for analysis, the remaining 70% were discarded due to a lack of sufficient FoW rating time points. Second, there was a positive correlation of $r = .61$ between FoW and Aha! self-report, if the analysis was done collapsed across participants. The second analysis which yielded the key finding of no correlation ($r = .08$) was run on the level of the 180 trials, but seemingly without taking into account the hierarchical structure of the data. An additional analysis could have clarified these contradicting patterns, for example by computing correlations on the trial level for each individual and then reporting the mean correlation coefficient across individuals, but this was not reported. Third, the data in Laukkonen and Tangen (2018) were collected as part of another experiment which used a manipulation before the problem solving task (Laukkonen & Tangen, 2017). This means that the sample was not a homogenous group but consisted of participants from a conflict inducing condition (presented with a switching Necker cube) as well as of participants from a no-conflict condition (two alternating cubes). It remains unclear how this may have impacted both the FoW ratings and the post-hoc Aha! self-reports.

Therefore, we conclude that both FoW ratings and Aha! self-reports are valid measures of the insight solution process, but with a different focus. Post-hoc Aha! ratings cannot answer the questions posed in the present study. In order to measure how participants' problem solving process unfolds across time and in response to manipulations, repeated FoW ratings were our method of choice.

However, the endeavour to directly compare the FoW and Aha! measures is certainly worthwhile, and it needs to be noted that the negative finding from the Laukkonen study (2018) is in accordance with another recent study that found no difference in FoW ratings between solutions for which an Aha! had been reported compared to those where no Aha! was reported (Hedne et al., 2016). The Hedne study used the task domain of magic tricks (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013, 2014b) and had participants classify their solutions into insight and non-insight trials by post-hoc Aha! self-reports as well as give a confidence rating for each solution. In addition, their solving process was tracked with FoW ratings in order to investigate the extent to which metacognitive feelings are predictive of "insight" versus "non-insight" solutions. Hedne et al. replicated previous findings (Danek et al., 2014b) that insight solutions were more likely to be correct than non-insight solutions, an effect that has also been found in task domains other than magic (Salvi, Bricolo, Kounios, Bowden, & Beeman, 2016; Threadgold, Marsh, & Ball, 2018; Webb, Little, & Cropper, 2016; see Danek & Salvi, 2018, for a discussion of this "accuracy effect"). Moreover, as in prior studies (Danek & Wiley, 2017), confidence predicted both the Aha! experience as well as solution correctness. Interestingly, the relationship between confidence and correctness was modulated by solution type, with a stronger relationship for solutions classified as non-insight solutions. Nevertheless, the Hedne study was not designed to answer the question whether FoW ratings are predictive of successful solving, because they did not analyse the warmth data with regard to solution correctness. Further, they had excluded unsolved trials from their analyses (because it does not make sense to obtain Aha! self-reports if no solution is provided). However, another study demonstrated higher FoW ratings for solvers of the 9-Dot problem, as compared to non-solvers (Chein et al., 2010).

One aim of the present study was to answer the question whether FoW ratings predict solving performance by comparing solvers (with correct solutions) with non-solvers (who gave no solution). We hypothesised to find differential patterns of FoW ratings, depending on solving performance: If solvers correctly assess their closeness to solution shortly before solving and non-solvers are aware of the fact that they do not have any solution available, there should be higher warmth ratings for solvers than for non-solvers (as found by Chein et al., 2010) shortly before solution.

In the present study, we used warmth ratings as a measure of impasse. We reasoned that if participants feel stuck, the impasse would be reflected in flat or decreasing FoW ratings, indicating that they do not feel they are coming any closer to the solution. Instead of assessing warmth 6 or 4 times each minute (resulting in up to 40 FoW ratings per problem) like in Metcalfe's original experiments (Metcalfe & Wiebe, 1987), we asked for ratings only four times (namely in each new section of the experiment, after new information was provided), so that the main problem-solving task would be less frequently disrupted. This is similar to the Hedne et al. study (2016) with only five FoW rating time points. The present study was designed to overcome a main limitation of both the Laukkonen and Tangen (2018) and Hedne et al. study (2016), namely that a large number of trials had to be discarded (around 70% in both) due to missing FoW ratings. This was achieved by using a fixed sequence of ratings (one after each step of the experimental procedure) which had to be given before a solution could be provided.

Leading problem solvers into an impasse

In the present study, we attempted to lead problem solvers into an impasse by constraining the search space via the depletion of possible solution ideas. In an elaboration of the RCT, Öllinger et al. (2014b) proposed that the search space before and after an impasse can be constrained. Since there exist many different methods and techniques for achieving a particular magic effect, the initial search space for magic tricks is quite large, with several false, implausible solutions available. As long as these false solution candidates are considered, participants still explore the search space and may not feel stuck. Therefore, to make reaching an impasse more likely, we took away these false solution candidates. Our manipulation was similar in nature to the concept of "experimentally induced fixation" of Smith and

Blankenship (1989) who induced mental fixation by presenting misleading cues in a rebus puzzle paradigm, leading to lower solution rates. The same decrease in solution rates was achieved by providing problem solvers with misleading associates while solving a remote associates test (Smith & Blankenship, 1991; Storm & Angello, 2010). Storm and Angello warned participants that the given response words would not work as viable solutions. Our approach was slightly different: shortly after presenting the problem, we provided participants with often proposed but incorrect solution candidates and revealed that these were not valid. If one of these typical invalid solution ideas had actually been considered by participants (which was likely because they had been taken from previous studies, see methods), a search for another, less trivial solution was prompted. This manipulation should trigger impasse in two ways. First, by making the search space smaller and thus faster to be exhausted. Second, by emphasising and spelling out two implausible solutions which was expected to inhibit the activation of other solution candidates (based on Smith, 2003; Smith, Ward, & Schumacher, 1993), and thereby to trigger mental fixation, directly leading to the onset of an impasse.

Cues as a means to overcome the impasse

According to Ohlsson, "the impasse is broken by seeing the problem in a new way, just as the Gestalt psychologists claimed" (Ohlsson, 1992, p. 12). For this purpose, we used two types of solution cues, verbal and pictorial ones. Our rationale was that once a state of impasse is reached, an appropriate cue might help to relax the key constraint that prevented a solution and thus trigger a representational change (as predicted by the representational change theory, Knoblich et al., 1999). We define an appropriate cue as non-trivial and implicit, one that does not give away the full solution (Luo & Knoblich, 2007), but instead offers new information that helps to restructure the problem representation. It was shown already 20 years ago (Davidson, 1995) that if the external solution-relevant information provided as a cue is too revealing, it leads to a shortcut in the search process, and results in an incremental, non insight-like FoW pattern. Therefore, instead of giving explicit cues, a more fine-tuned method was needed to trigger constraint relaxation.

Previously, both with overt (Grant & Spivey, 2003) and covert (Thomas & Lleras, 2009) cues, it was possible to trigger more solutions in the cueing condition, as compared to baseline conditions. Also, priming (Slepian, Weisbuch, Rutchick, Newman, & Ambady, 2010), breaks (Beeftink et al., 2008) and sleeping (Wagner, Gais, Haider, Verleger, & Born, 2004, but see Schönauer et al., 2018, for a different finding) were found to influence the occurrence of insight. Moreover, beyond just increasing solution rates, Öllinger et al. (2014b) could show that it is possible to trigger constraint relaxation by cueing. They implemented cues that separately manipulated each source of difficulty (Kershaw & Ohlsson, 2004) in the 9-dot problem. Further, it was shown that it is possible to overcome an impasse with external help (e.g. Öllinger, Jones, Faber, & Knoblich, 2013; Öllinger, Jones, & Knoblich, 2014a), but it remains unknown what kind of cues lead reliably to the solution in the domain of magic.

In contrast to most studies that use only one type of cue to aid the solving process, the aim of the present study was to compare two types of cues (verbal and pictorial) with a baseline condition (no cue). Ormerod et al. (2002) utilised both visual and verbal cues to increase solution rates in different versions of the eight-coin problem. If participants could not find the solution on their own, they received two consecutive verbal hints. Further, one group was provided with an initial visual hint pointing to the third dimension (stacked coins). Surprisingly, providing such an explicit visual hint did not lead to higher solution rates. However, another, higher-powered study did show the efficacy of a similar visual cue (overlapping coins) for this problem (Öllinger et al., 2013). Another study (Pedone, Hummel, & Holyoak, 2001) used visual hints (Experiment 2), then visual and verbal hints combined (Experiment 3) to enhance analogical problem-solving performance. This study found that though verbal hints facilitated the recall of source analogues, visual hints were still more successful in enhancing solution rates. In the context of training visuospatial processing skills in children, Chabani and Hommel (2014) compared the efficacy of visual as well as verbal and visual cues combined on a visuospatial task. The visual cues turned out to be more effective than the combined cues. The authors argued this result might be due to visual processing being more closely related to a spatial task than verbal processing. In contrast to cues from the same modality, different-modality cues seem to entail modality switching costs (Connell & Lynott, 2011). In sum, these studies show that participants are able to extract information from different types of helpful cues and use them for solving a problem. However, none of these studies compared two types of cues separately in a within-subject design so it remains unclear which type of cue is more effective. In the present study, using visual stimulus material (video clips of magic tricks, see below), we compared verbal vs. visual (pictorial) cues and expected a higher efficacy for pictorial ones because they are presented in the same modality.

We developed a set of cues as additions to a recent paradigm (Danek et al., 2013), which is bridging the gap between the science of magic (Kuhn, Amlani, & Rensink, 2008) and the growing demand for a set of similar and relatively quickly solvable insight problems (Bowden et al., 2005). This paradigm uses magic tricks as a problem-solving task, asking participants to find out how each trick was carried out. As outlined in Danek et al. (2014b), since the magician uses his skills to subtly manipulate observers' perception (e.g. their assumptions about the objects used), watching a magic trick is quite likely to trigger a biased, false problem representation. To be able to see through the magic trick, participants must overcome their initial view of the problem through a representational change, for example by relaxing their false assumptions. Thus, the task domain of magic offers the possibility to investigate the phenomenon of insight. Participants were assumed to get into a state of impasse as a stage of the insight problem-solving process since after one presentation of a magic trick it is hardly possible to reach insight and solve the problem (based on Danek et al., 2014b).

In particular, this selected set of magic tricks shares one important aspect: there is only one key element that is crucial for the solution. The cues were developed to point towards this detail in order to guide out of the impasse. This entails either presenting a "structural" analogy (Dahl & Moreau, 2002) with the key move or actually highlighting the key detail in a way which still does not make the full solution immediately accessible. For a smaller subset of 12 tricks, verbal cues had already been successfully implemented to increase solution rates by about 12% (Danek et al., 2014b).

Aim and hypotheses

The main aim of this study was to measure how participants' problem solving process unfolds across time and in response to two manipulations. We tried to lead problem solvers into a state of impasse by constraining the search space, and then out of the impasse with the help of solution cues. First, we predicted that taking away two implausible, trivial solution candidates would lead to impasse, indicated by non-increasing FoW ratings. Second, giving cues should lead out of the impasse, indicated by increasing FoW ratings. With regard to the predictive value of FoW ratings on solution correctness, we hypothesised that correct solvers would give higher FoW ratings than non-solvers who could not come up with any solution. Further, we hypothesised that solution rates in both cueing conditions would be higher than in the baseline condition. Specifically, we expected solution rates to be higher with pictorial cues than with verbal cues because they matched the modality of the visual stimuli.

Methods

Participants

Participants were recruited from the Eötvös University and the Budapest University of Technology and Economics, Hungary, as well as from the

Ludwig-Maximilians-Universität, Munich, Germany. 47 healthy participants aged between 18 and 29 years ($M = 23.38$, $SD = 2.21$) volunteered to complete the experiment, each of them signed an informed consent in accordance with the Declaration of Helsinki ("WMA's Declaration of Helsinki," 1964) and its later amendments. This study was carried out following the recommendations of the Institutional Review Board of the Ludwig-Maximilians-Universität, Munich. All volunteers were international students who had graduated at least from high school, 21 of them were men (44.6%). All participants reported normal or corrected-to-normal visual acuity and were compensated with 8€/hour for their participation.

Stimuli and apparatus

Stimuli were presented on a 15.6-inch monitor screen having a 60 Hz refresh rate and a resolution of 1366x768 pixels. Participants were seated 50 cm from the screen. The experiment was coded using Psychopy software, v1.74 (Peirce, 2007).

A set of 33 magic tricks which had been recorded with professional magician Thomas Fraps (Abbott, 2005), was used as problem solving task. Participants watched each magic trick two times with the task of finding out the secret method used by the magician. From the original set of 34 video clips of short magic tricks (each consisting of only one effect and one method), one trick was excluded due to the inability of cueing it.

Procedure

After reading the instructions, subjects completed a practice block of three trials, followed by the main experiment consisting of 33 trials. The 33 magic tricks were randomly distributed across the three cueing conditions, so that each subject saw 11 tricks in each condition (verbal, pictorial or no cue). The task in each trial was to find out the secret method behind the trick, i.e. to explain how the magician achieved the magic effect. No feedback was given about solution correctness. The trial sequence as depicted in Figure 1 was identical for all participants, with the exception of the cue which depended on the randomly selected cueing condition for that trick.

First, a 500 ms fixation cross appeared in the centre of the screen. Subsequently, the magic trick video clip was shown for the first time. Then participants had a self-paced thinking time to process what they had just seen and to try to solve the problem without any distractions. A black screen was displayed until they pressed a button, which led them to their first FoW rating. Subsequently, two implausible, trivial solutions of the trick were displayed for 10 seconds, complemented with the statement that these were

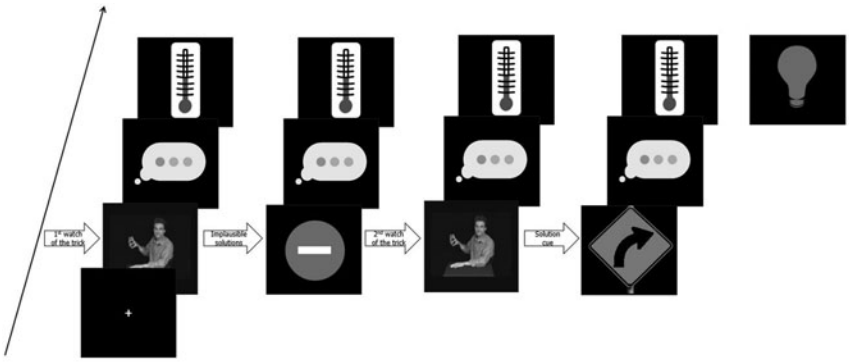

Figure 1. Schematic sequence of one trial. Symbols used: Magician = presentation of the magic clip, typing sign = thinking time, thermometer = FoW rating, no entry sign = implausible solutions, turn right sign = solution cue, lightbulb = solution prompt.

not the correct solutions. Next, the "thinking time – FoW rating" sequence was repeated. Then participants watched the magic trick for the second time after which they went through the "thinking time – FoW rating" sequence again. This was followed by the cue manipulation: either a verbal, pictorial or no cue was presented, depending on the randomly selected cue condition. Next, the final self-paced thinking time took place. After that, subjects were asked whether they knew the solution of the trick. If not, they gave their fourth FoW rating. If they had a solution idea, they had to indicate how certain they were about it and then were asked to type in their answer. Since this certainty rating was given on the same scale from 1-8 and was another meta-cognitive appraisal of subjective solving performance, we used it as the fourth FoW rating also for participants who provided a solution idea. Please note it was not possible to skip parts of the trial if a solution idea occurred earlier, participants were guided through the entire trial sequence each time. All FoW ratings were made by button presses, and solution ideas were typed in using the keyboard. As soon as the solution idea got submitted, the next trial started. Finally, participants took part in a short debriefing survey including a question about familiarity by which they had to indicate whether the solution of a trick had already been known to them before the experiment. The testing session lasted between 70 and 100 minutes including a 5 minutes break after completing half of the trials.

Design

The experiment was set up in a within-subjects design, with the cueing condition (three levels) as within-subjects factor. The 33 tricks were divided randomly into the three cueing conditions and each participant received an independently randomised sequence of the 33 tricks. Each trick was

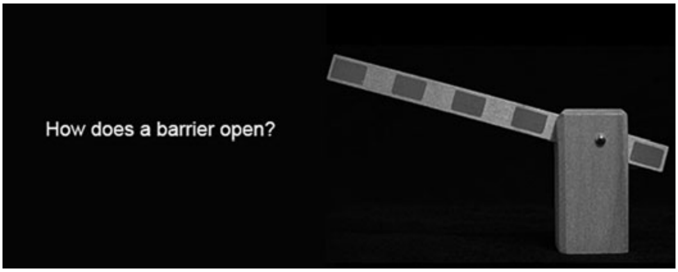

Figure 2. Verbal (on the left) and pictorial (on the right) cue for the "Matchstick" trick. In the trick clip, the magician holds one matchstick between the forefinger and thumb of each hand and then quickly moves one matchstick through the other one, seemingly penetrating the matchstick in the middle without breaking or damaging it. The solution is that the magician quickly raises his forefinger to which the vertical matchstick is glued to let the other matchstick pass through.

presented together with the corresponding implausible solutions and the corresponding cue (either no cue, a verbal cue, or a pictorial cue).

Cues

To test the efficiency of different types of solution cues, a verbal and a pictorial cue was created for each trick. Both were designed to trigger a representational change by hinting at the critical solution detail, but did not give away the full solution. Verbal cues consisted of a sentence which guided the attention to the crucial problem element. Pictorial cues consisted of an analogy of the crucial element (for an example see Figure 2). Both types of cues were extensively pre-tested, and the final selection was based on their efficiency in enhancing the solution rate. A German version of 12 of the verbal cues was already shown to be efficient in Experiment 2 of a previous study (Danek et al., 2014b), these were translated to English. The remaining 21 verbal cues as well as all 33 pictorial cues were newly developed. There was also a third condition (baseline) in which no cue (just a grey rectangular shape) was presented.

Implausible solutions

To accelerate the onset of an impasse, all participants were presented with two sentences describing two solutions which were not viable. These "implausible solutions" were composed of actual implausible solutions given by participants of previous experiments. For example, in the trick "Matchstick" – shown below – the two sentences were: "The match has no opening which closes again." and "The matches are not broken."

FoW

Participants were asked "How close are you to the solution of the trick?" and had to select a value on a proportional scale with numbers from 1 to 8.

Participants gave ratings on four different occasions (see Figure 1), namely every time they had received new information. The first FoW rating appeared after watching the trick for the first time; the second was given after the presentation of the implausible solutions; the third one after watching the trick for the second time; and finally, the fourth FoW had to be made after the presentation of the solution cue.

Thinking times

We aimed to personalise the problem-solving experience to favour each participant's solving style and to avoid frustration. Therefore, participants had dynamic thinking times (blank screen) instead of fixed times which had turned out to be rather frustrating for participants in pilot studies. There was no time-out, so that participants could chose for themselves how long they thought about the problem during each thinking time window. To avoid disruptions of the solving process, we inserted the FoW questions after the thinking times (see Figure 1), so that both the viewing of the magic trick as well as the thinking time was completed without interruption.

Data analysis

To evaluate solution accuracy, a solution catalogue derived from several prior studies which additionally had been verified by the magician was employed to code the answers into four categories. The first category comprises the real solution (i.e. the method that the magician used) and the second comprises alternative, but plausible solutions, while the third category consists of partial or implausible solutions, and the fourth comprises trials without any solution attempts (no solution given). Two independent raters coded the solution texts into the four solution categories. The inter-rater reliability analysis resulted in a Kappa value of 0.938, indicating a strong agreement between the raters which is not due to chance ($p <$ .001). If a participant was already familiar with a trick, their response was excluded from analysis and handled as a missing value. 33 magic tricks were shown to 47 participants, resulting in a total number of 1551 trials. All in all, $n = 97$ trials (6.25% of the data) were lost due to technical problems, $n = 14$ trials (0.9% of the data) were excluded since participants knew the trick already. This resulted in a dataset of $n = 1440$ trials. One trick was not solved by any participant in the present sample, and was thus excluded ($n = 45$ trials) from further analysis. The analysed sample consists of $n = 1395$ trials.

Dependent variables were solution rates and FoW ratings. Solution rate means the percentage of correctly solved tricks for each participant averaged over in each condition. For each participant, his/her percentage of solved tricks was computed by dividing all solved tricks by all presented tricks (e.g., if 11 tricks were solved from the 33 presented ones, the solution rate of the participant was 33.3%). If tricks were excluded due to the reasons mentioned above, the total number of solved tricks was divided by the accordingly adjusted number of presented tricks.

The Statistical Package for Social Sciences (SPSS) version 20, R statistical software programme (R Development Core Team, 2016) and Microsoft Office Excel 2010 software were used to aggregate and analyse the data. For the ANOVA, degrees of freedom were corrected with Greenhouse-Geisser estimates if the assumption of sphericity was violated ($\varepsilon \leq .8$).

Results

Solution rates as a function of cue condition

Participants came up with solution ideas in 63.6% of the 1395 trials; the overall percentage of correct solutions is 38.1%. For a detailed overview see Figure 3. Trials in which a partial or implausible solution was given were excluded from all further analyses, because the main comparisons of interest were between solvers (correct or plausible alternative solution) and

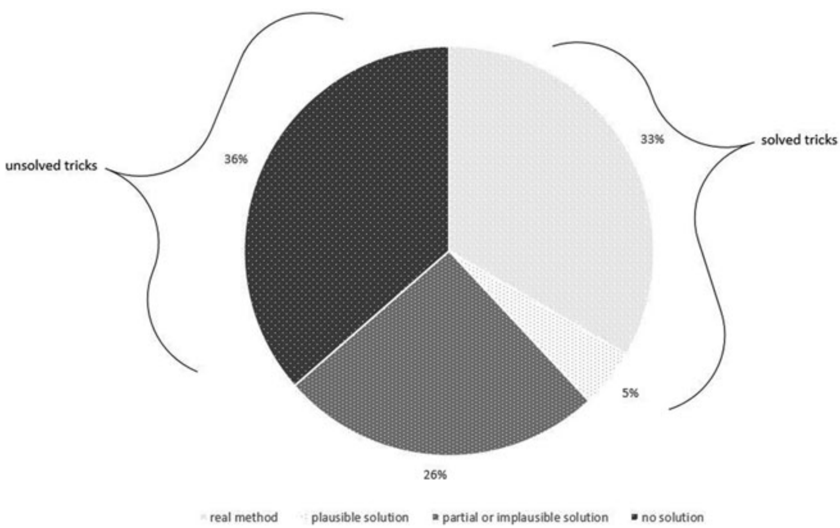

Figure 3. Coded solutions per category. The light grey patterned area depicts the percentage of solutions in which participants found out the real method behind the trick; the white dotted area stands for alternative, but plausible solution suggestions. The sum of these two forms the 'solved tricks' category (38%). The dark grey dotted area depicts partial or implausible solution ideas. In the remaining black patterned area, the proportion of no solutions is shown.

non-solvers (no solution given). Since participants never received feedback about their solution, they were not aware that their implausible solutions were incorrect.

The solution rates collapsed by the 32 tricks across all participants ranged between 10% ("Restoration" was the most difficult trick) and 83% ("Cigarette" was the easiest trick) with an average solution rate of 40% (*SD* = 14.6%). The solution rates collapsed by the 47 participants varied between 2 and 21 from the 32 tricks ($M = 11.72$, $SD = 4.26$).

Considering differences in solution rates between the three cue conditions, the largest number of tricks were solved correctly with the help of a pictorial cue (49.2% of all trials of this condition), a smaller number with a verbal cue (38.8%) and even less (28.6%) in the baseline condition without any cues, as illustrated in Figure 4.

Conducting a one-way repeated measures ANOVA with the solution rate in each condition as dependent variable, the expected main effect of cue condition was found [$F(2,72) = 14.4$, $p < .001$, $\eta^2_{partial} = .239$]. Subsequent analyses (paired-samples t-tests) showed that the outcomes differed significantly between presenting pictorial or no cues [$t(46) = 4.4$, $p < .001$, CI

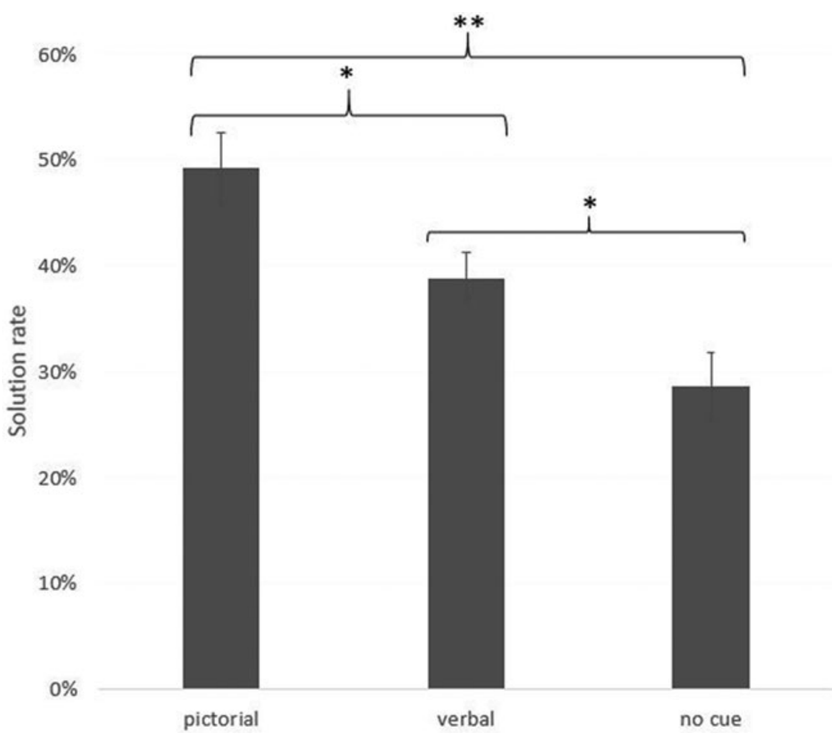

Figure 4. solution rates for the three cue conditions. $^{**}p < .001$, $^{*}p < .01$. Error bars denote standard errors of the mean.

95% [0.11, 0.3]]. The difference in the solution rates between the verbal and no cue conditions [t(46) $=$ 2.78, $p < .01$, CI 95% [0.03, 0.18]] and between the pictorial and verbal cue conditions were also significant [t(46) $=$ 3.5, $p < .01$, CI 95% [0.04, 0.16]].

FoW ratings as a function of rating time points

FoW ratings were analysed on the level of observations. The data were analysed on various levels of differentiation, adding more factors to the analyses in each step: rating time points, then solving performance, then cues.

First, we tested whether the FoW ratings differed across rating time points. We carried out a linear mixed-effects regression (LMER) analysis using R (R Development Core Team, 2016) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2014). LMER is different from the mixed ANOVA analysis since it does not require data aggregation in order to assess the variance related to fixed and random factors (see Baayen, Davidson, & Bates, 2008).

To test the main effect of time points, a likelihood ratio test was executed. This works by comparing a null model (model without rating time points) to a full model (model with rating time points). The result of the comparison was significant ($\chi^2_{(3)} = 137.36$, $p < .001$).

Multiple comparison tests were carried out in order to know which levels of rating time points were significantly different from each other. The Tukey posthoc test was used to adjust the p-values influenced by multiple comparisons. From the comparison p-values, only FoW_1 and FoW_2 were not significantly different from each other ($p > .05$), each of the other comparisons yielded a significant result ($p < .001$). This means, overall, the first two FoW ratings did not differ from each other, but all others did. This is a first indication that initially, participants felt they were not coming closer to a solution. However, to gain a clearer picture it is necessary to consider the FoW ratings for solved and unsolved trials separately.

FoW ratings as a function of solving performance

In order to investigate whether solvers and non-solvers showed differential patterns of FoW ratings, we ran another LMER analysis. As fixed effects, we entered solution of tricks (solved/unsolved) and time points coupled with their interaction, while as random effects, we entered intercept for subjects to control for variability due to multiple responses from each subject. The p-value for the effect of solution of tricks was obtained using likelihood ratio tests of the full model (with solution of tricks included) against the null model (without solution of tricks included). Solution of tricks affected

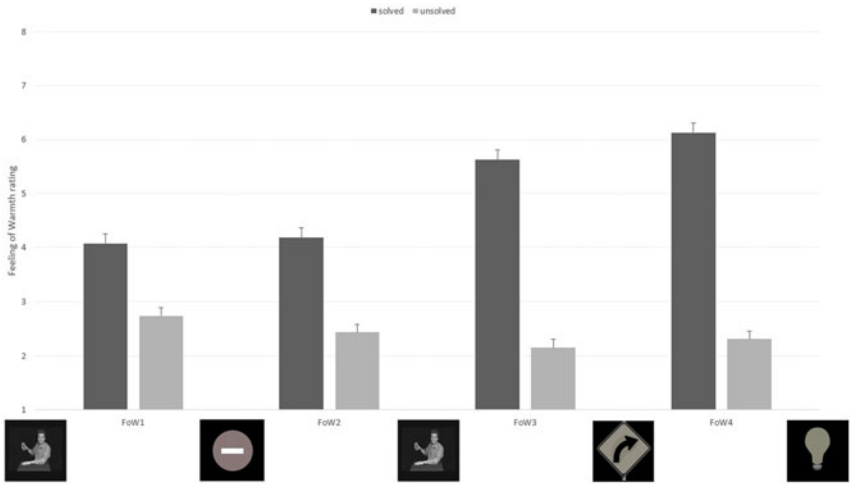

Figure 5. Mean FoW ratings (on a scale from 1 to 8) as a function of solved and unsolved trials. Note that each experimental step (depicted as icons) was followed by a self-paced thinking time, compare Figure 2. Solved trials had initially flat ratings, with a subsequent increase in warmth. Unsolved trials had initially decreasing ratings, which ended flat. Error bars denote standard errors of the mean. Note that the cue manipulation contained a helpful cue in only two-third of trials.

FoW ratings significantly. Figure 5 shows that the FoW ratings at any given time point differed between trials that would later be solved as compared to trials that would not be solved. Also, when ratings were accumulated across all 4 timepoints, the mean of FoW ratings by the solved trials ($M = 5.07$, $SD = 1.83$) was significantly higher than those by the unsolved trials ($M = 2.26$, $SD = 1.21$) indicating that participants could correctly predict their subsequent performance.

Our prediction was that taking away two implausible, trivial solutions would lead to impasse (expressed in non-increasing FoW ratings). The finding of flat (solved trials) or even decreasing (non-solved trials) warmth ratings after the implausible solution manipulation (plus subsequent thinking time) supports this impasse hypothesis (see Table 1 and Figure 5). Considering only solved trials, there is an initial plateau, followed by a clear increase in the FoW ratings. This means, after a period of impasse with flat, non-increasing ratings between FoW_1 and FoW_2, participants felt that they got increasingly closer to the solution. This increase was significant between FoW_2 and FoW_3 (after the second viewing of the trick) as well as between FoW_3 and FoW_4 (after the cue). This pattern is different for unsolved trials, with a significant cutback in the ratings from FoW_1 to FoW_2 (after implausible solutions were taken away) indicating that participants were guided "deeper" into the impasse by this manipulation. The decrease continued, as FoW_3 was also significantly lower than FoW_2. With the cued and uncued trials mixed together, this analysis has shown no further

Table 1. Multiple comparisons parameter estimates across the FoW ratings in solved and unsolved trials.

		Mean difference score	*SEM*	*t*	*df*	*p*
Solved trials	FoW_1-FoW_2	−.01	0.11	−.93	2124.1	.788
	FoW_2-FoW_3	−1.44	0.11	−12.78	2124.1	.000
	FoW_3-FoW_4	−.5	0.11	−4.5	2124.1	.000
Unsolved trials	FoW_1-FoW_2	.3	.08	3.63	2031.17	.002
	FoW_2-FoW_3	.28	.08	3.38	2031.17	.004
	FoW_3-FoW_4	−.16	.08	−1.87	2031.17	.241

Note. Negative difference scores indicate an increase.

change in warmth after the cue (no difference between FoW_3 and FoW_4). Basically, if participants were not able to solve the trick, their ratings stayed on a very low level, see Figure 5.

This analysis included all three types of cues, so a third of the cues was not helpful (no cue). To further disentangle the interplay between impasse, solving performance and the influence of cue condition, another analysis was conducted.

FoW ratings as a function of cue condition

To inspect the effect of the presence of solution cues on FoW ratings, an LMER analysis was executed. For this analysis, both the pictorial and verbal cue conditions were aggregated into a "cue" condition and compared with the "no cue" condition. Both solved and unsolved trials were included. As fixed effects, we entered presence of cues and time points along with their interaction, while as random effects, we again entered intercept for subjects. The effect of cue condition was tested by running likelihood ratio tests, which compare the full model (including cues) to the null model (not including cues). The main effect of the presence of cues was significant.

Post-hoc comparisons revealed that the implementation of the helpful cue led to higher warmth ratings. Independent of whether the trick was ultimately solved, FoW_4 was higher in the cue condition ($M = 4.61$, $SD = 2.63$) than in the no cue condition ($M = 3.42$, $SD = 2.54$). This means that receiving a (verbal or pictorial) cue let participants feel closer to solution in FoW_4 as compared to receiving no cue which left them unaffected, see Table 2 (results in bold). This finding is in accordance with having higher solution rates for cued trials vs. non-cued trials. As expected, the other interaction parameter estimates terms were not significantly different from zero ($|$t values$| < 1.96$), as the cue had not yet been introduced during the earlier rating time points. This means, during the first three rating time points, FoW ratings changed in the same manner (for the values, see Table 2), independent of cue condition, namely no change between FoW_1 and FoW_2 and a significant increase between FoW_2 and FoW_3.

Table 2. Multiple comparisons parameter estimates across the FoW ratings in cued and uncued trials.

		Mean difference score	*SEM*	*t*	*df*	*p*
Cue	FoW_1-FoW_2	.09	.02	.83	2838.02	.839
	FoW_2-FoW_3	−.61	.02	−5.55	2838.02	.000
	FoW_3-FoW_4	**−.58**	**.02**	**−5.22**	**2838.02**	**.000**
No cue	FoW_1-FoW_2	.1	.15	.63	1347.37	.921
	FoW_2-FoW_3	−.58	.15	−3.77	1347.37	.001
	FoW_3-FoW_4	**.17**	**.15**	**1.08**	**1347.37**	**.700**

Note. Negative difference scores indicate an increase. The cue was given between FoW_3 and FoW_4, so this result is depicted in bold.

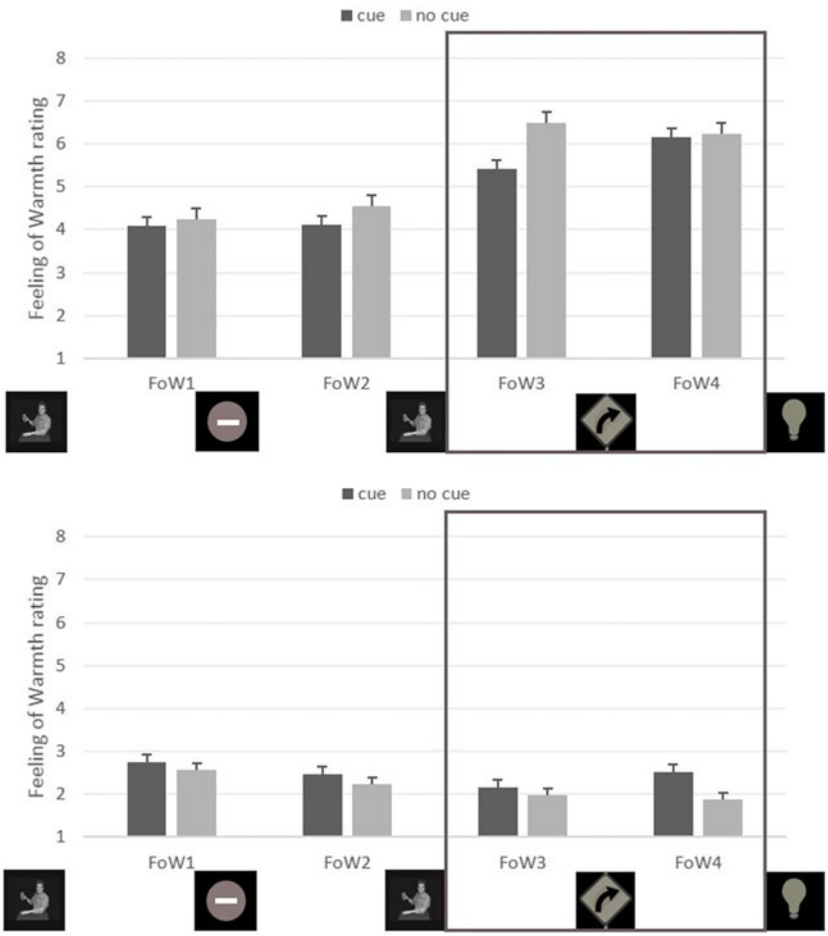

Figure 6. Mean FoW ratings for solved (upper panel) and unsolved (lower panel) trials when a (pictorial or verbal) cue was presented (dark grey bars) as opposed to the baseline condition (no cue presented, light grey bars). The cue manipulation took place between FoW_3 and FoW_4 and is symbolised by the "turn right" sign. Error bars denote standard errors of the mean.

To illustrate the differences between solvers and non-solvers, Figure 6 depicts the ratings again separately for solved and unsolved trials. Solvers felt significantly closer to the solution after receiving a helpful cue, whereas if no cue was presented, their ratings did not change. In case of the non-solvers, the same pattern was observed but in lower magnitude.

Discussion

Addressing the theoretically postulated impasse-insight sequence, the present study investigated how problem solvers get into and out of a state of impasse. The likelihood of reaching an impasse was increased by taking away implausible, trivial solution ideas. We also examined which types of cues were the most efficient in guiding problem solvers out of the impasse.

As expected, there was a main effect of cue condition, showing that both pictorial and verbal cues were more helpful in guiding to solution than a baseline condition with no cues. Since the cues were specifically designed to relax constraints that initially prevented a solution, their efficacy supports the idea of constraint relaxation as one possible mechanism to overcome impasse (Knoblich et al., 1999; Ohlsson, 1992). It is noteworthy that none of the cue types led to performance at ceiling (solution rates after cues $< 50\%$), which indicates that the cues did not simply give the solution away, but that they only guided participants towards the solution by allowing them to relax crucial constraints. Specifically, verbal cues led to a 10% increase in solution rates which is comparable to the 12% increase obtained in a previous study using a subset of the same verbal cues (Danek et al., 2014b). Pictorial cues led to a 20% increase, and, as expected, were significantly more efficient than verbal ones, although both cue types basically contained the same information. One could argue that pictorial cues have an advantage over verbal cues because they might be less ambiguous and leave fewer degrees of freedom for interpretation than verbal ones. However, the verbal cues used here were rather specific. A more likely explanation seems that pictures are more similar in modality to the video clips than written sentences. A study by Grant and Spivey (2003) supports this explanation: Duncker's radiation problem (Duncker, 1945) was presented either in static form as a diagram with no animation (baseline), or in animated form where attention was drawn to a pulsating problem element (either critical or non-critical for solution). The critical-element animation condition led to twice as high solution rates than either the static or non-critical animation condition. This experiment demonstrated a very high cue efficacy when the modality of the cue (visual) was matched with the modality of the problem presentation (also visual, namely a diagram). Sas et al. (2010) investigated what types of visual cues could serve as helpful analogies to solve a visual insight problem, the eight-coin problem. Although a comparison with previous data (Ormerod et al., 2002) has shown that the implementation of verbal cues led to higher solution rates overall, visual cues had a beneficial effect on task performance, too. Sas et al. suggested that visual cues are best introduced after an impasse has already been reached and that visual cues can be processed more fluently than verbal ones due to the superiority of the human image-processing capacity over the linguistic one.

Overall cue efficacy was also reflected in the warmth ratings. There was a clear increase from FoW_3 to FoW_4, indicating that participants felt closer to the solution after a helpful cue, as opposed to no change in the baseline condition, as shown in Figure 6. This effect was observed independent from whether participants were ultimately able to solve the problem, but it took place on two different levels of warmth, i.e. on a higher level for later solvers. Thus, the cue manipulation worked as intended, helping problem solvers out of their impasse (reflected in the FoW increase) and, at least subjectively, guiding them closer to solution. This shows that constraints can be specifically targeted by appropriate cues, enabling problem solvers to relax these constraints and consequently overcome the impasse, at least in the case of solvers. Although non-solvers' warmth ratings stayed on a very low level, those who received a helpful cue could profit from it (slight, but significant increase from FoW_3 to FoW_4). Still, it seems not justified to conclude that non-solvers actually got out of the impasse, because they still felt very far away from a solution and in fact were not able to ultimately solve the problem, but perhaps they gained a vague solution idea from the helpful cue.

The implausible solution manipulation led to flat (solved trials) or even decreasing (non-solved trials) warmth ratings and thus seemed successful in leading problem solvers into impasse. After the manipulation, participants could not profit from the subsequent thinking time although it was self-paced and theoretically unlimited. Figure 5 shows that over the course of the entire solution process, solvers had initially flat ratings, followed by an increase in warmth and that non-solvers had initially decreasing ratings, which ended flat. These differences imply that non-solvers stayed in a state of impasse while solvers got out of the impasse which is also reflected in their solution success. After watching the trick for the first time, solvers already differed from non-solvers in their initial assessment of closeness to solution, giving higher FoW_1 ratings than non-solvers. However, both ratings were at a rather low level, indicating that also solvers did not have a full solution idea yet.

The following speculative interpretation fits the observed pattern of solvers: After the first viewing plus thinking time, solvers might have had an intuition or hunch, but no clear hypothesis about how the trick could work (rather low FoW_1). This intuition was not related to any of the two implausible solution ideas, and thus they neither became discouraged nor more confident when these were taken away (FoW_2 on same level as FoW_1). But it can be assumed that they were still in an impasse because also the second self-paced thinking time did not help them to feel any closer to solution. However, the second viewing of the trick together with the third thinking time allowed them to feel closer (steep increase from FoW_2 to FoW_3). The second viewing offered an enrichment of the existing problem representation, perhaps allowing them to elaborate their initial intuition and to form a hypothesis. The present design does not allow to answer the question whether or when exactly solvers experienced insight, but since FoW_3 ratings were still not at ceiling, we conclude that solvers were still not completely sure about their solution idea. They needed another helpful manipulation, the cue, to further increase their FoW rating (FoW_4). The specific influence of cues has already been discussed. That FoW ratings never reached the maximum of the scale, not even for solvers, probably reflects the difficulty and unfamiliarity of the task of seeing through a magic trick, and also perhaps a tendency to avoid extreme values.

We interpret non-solvers' problem-solving activity as depicted in Figure 5 as follows: the decrease from FoW_1 to FoW_2 indicates that non-solvers became discouraged or confused from taking away the implausible solutions, probably because they had entertained one of these ideas. This shows that the experimental manipulation for guiding people into a state of impasse worked: by pointing out that two typical trivial responses were not the correct ones, we "took away" some first solution ideas, perhaps even triggering a fixation on them. A caveat here is that the present design does not allow to directly verify if participants were indeed thinking of one of the two implausible solution ideas. However, this appears quite likely because these solution ideas had been repeatedly and consistently brought forward by many participants in three prior studies using the same problems. They seem to be the standard answers that first come to mind, although upon closer inspection, they turn out to be implausible. After this manipulation, non-solvers could not profit from the second viewing of the magic trick and the subsequent thinking time at all (in contrast to solvers), but instead felt even further away from the solution than before (decrease from FoW_2 to FoW_3). This finding again speaks for the interpretation that they became fixated by the implausible solution ideas. They could also not profit from the cue anymore (see previous section). Taken together, the impasse that non-solvers experienced and which they failed to get out of was clearly reflected in the low and rather flat line of FoW ratings. This finding is in accordance with recent data from another task domain (German

CRAT problems) where non-solvers also had a flat line in FoW ratings (Kizilirmak et al., 2018). In sum, the present data support our interpretation that the manipulations have guided both solvers and non-solvers into an impasse, and guided solvers out of the impasse again.

As expected, the FoW ratings showed distinct profiles for solved and unsolved trials. Solvers gave higher warmth ratings than non-solvers at each time point of the problem-solving process. Chein et al. (2010) also found higher warmth ratings for solvers, but only in the last 15 seconds before solution, not earlier (analyzing a time window of the last 90 seconds before solving the 9-dot problem). In our study, already the first exposure to the magic trick together with some thinking time seems to have offered problem solvers a sufficient reference base for predicting the subjective difficulty of that task, because there was already a small gap between solvers and non-solvers at the first FoW rating and this gap only grew wider across time (see Figure 5). That this difference already emerged after the first encounter with the problem suggests that participants had correct initial (and subsequent) metacognitions about their ability to solve it. This is in accordance with another study (Žauhar, Bajšanski, & Domijan, 2016) which found a positive relationship between FoW ratings and accuracy on a rule-based category learning task. However, that non-solvers' warmth ratings stayed on a low level is in contrast with research showing people often make overconfidence errors in several cognitive domains (see Metcalfe, 1998, for a review). In particular, Metcalfe showed that high FoW ratings shortly before an answer predicted that the answer to classical insight problems would be incorrect (Metcalfe, 1986). In contrast to the Metcalfe study, non-solvers in our study were people who did not provide any solution at all (instead of people who had suggested an incorrect solution) which could be a reason for the differing findings. Further, it could be argued that the specific task domain of magic tricks does not lead people into overconfidence because prior experience with magic tells that it is nearly impossible to see through a magic trick. Finally, the present result of non-solvers having flat FoW ratings corroborates recent findings from Kizilirmak et al. (2018).

The present data are consistent with key assumptions of the representational change theory of insight (Knoblich et al., 1999; Ohlsson, 1992): First, the occurrence of impasse. The discontinuity of a problem solver feeling stuck in an impasse has no place in the business-as-usual view on problem solving, which conceptualises insight as the result of analytic search processes that unfold continuously (e.g. Fleck & Weisberg, 2004; MacGregor, Ormerod, & Chronicle, 2001). In contrast, the present non-increasing warmth ratings for both later solvers and non-solvers between FoW_1 and FoW_2 indicate a phase of no continuous subjective progress during two self-paced thinking times where participants could search for a solution.

Second, the warmth pattern of solvers includes one steep increase between FoW_2 and FoW_3 which is also more in line with the theory of sudden representational change than with the business-as-usual view of insight problem solving as an incremental search (as already argued by Metcalfe & Wiebe, 1987). Third, the high efficacy of the cues which had been specifically designed to offer new information which contradicted participants' initial, constrained problem representation, and thus to trigger a change in representation, is in accordance with the idea of constraint relaxation as one possible mechanism to overcome impasse (Knoblich et al., 1999; Ohlsson, 1992).

Several limitations of the present work need to be discussed. The focus of our study was on the onset/offset of impasse and the influence of cues, but not on the actual moment of insight. Thus, we did not track the actual time point of when participants reached a solution but forced participants to go through the entire sequence of two magic trick viewings plus four FoW ratings. While this procedure means that warmth ratings before and after the actual solution time point could not be differentiated, but were analysed together, we nevertheless decided for this procedure because it has two main advantages: Each participant contributed the same data to each FoW rating time point, instead of solvers dropping out earlier and not contributing to later FoW ratings anymore (an issue already raised by Weisberg, 1992). In addition, we could analyse all the rating data in contrast to previous studies which had to discard large proportions of trials (about 70%) due to missing ratings (Hedne et al., 2016; Laukkonen & Tangen, 2018).

We also had to make a decision about whether to measure the subjective occurrence of insight in this study in any additional way besides the FoW ratings. Since it has repeatedly been shown that magic tricks trigger strong Aha! experiences (e.g. Danek et al., 2014a; Hedne et al., 2016) and since this was not the point of the present study, we deliberately refrained from asking participants about the quality of their solution (i.e. whether they solved with or without insight, whether they had an Aha! experience) to avoid any bias on our dependent variable (FoW ratings) that could be introduced by describing a priori how an insight feels like (e.g. sudden, pleasant). Thus, the present data cannot speak for or against the viewpoint that problem solvers experienced insight during solving, only that they eventually solved the task or not. However, whether a moment of insight occurred or not could still be mirrored in changes of FoW ratings across time. Although we analyzed group means and not the raw frequencies when comparing solvers with non-solvers, we found a similar step towards high ratings as Metcalfe and Wiebe (1987) did. The sharp increase in solvers' FoW ratings between FoW_2 and FoW_3 (before any cue was provided, compare Figure 5) could be interpreted as a sudden insight, but the same could be said about the increase between FoW_3 and FoW_4. Without any further information about whether or when participants experienced an insight (Aha! experience) or not, it seems not justified to draw any conclusions from this. In a similar vein, despite using FoW ratings, the present study was not in a position to answer the longstanding question about how sudden or gradual the solution process usually unfolds. The reason for this is that although the dynamics of the solution process was tracked across time, our two manipulations may have altered the naturally occurring dynamics. There is some evidence that cues may not only influence solution rates, but also lead to a thinking process that is perceived as more incremental (Bowden, 1997; Cushen & Wiley, 2012). The same argument limits the comparability of the present results with other studies using FoW ratings in which the problem-solving process is typically not manipulated. Basically, participants in our study experienced "cued or induced insight," as compared to purely self-generated insight without any outside help.

It is also important to note that for participants who provided a solution, the fourth FoW rating was in truth a confidence rating that used the same format of the FoW scale (compare methods). Although this is also a metacognitive appraisal of subjective solving performance and the two ratings are clearly related, they do not measure identical constructs. A FoW rating measures subjective progress towards a potential solution and a confidence rating measures subjective certainty about a found solution. However, we believe that this did not impact the interpretability of our findings because the clear difference between solved (high FoW) and unsolved trials (low FoW) was already apparent at FoW_3 (which was a true warmth rating for everyone) and did not change much at FoW_4 (which was a confidence rating for solvers), as can be seen in Figure 5.

Finally, although the cues clearly increased overall solution rates, they may have done so differently for different tricks. Magic tricks are very particular stimuli, and the present set of 33 tricks could be argued to be more heterogeneous than, for example, a set of verbal insight puzzles (although these also refer to different situations and draw on different knowledge elements). Some tricks remained very difficult to solve even with a cue (the most extreme example being the trick with 0% solution rate both with and without cue which we excluded from the analysis). Thus, as a caveat we need to point out that it is possible that differences in the FoW patterns between solved and unsolved trials may also reflect differences in the problems and not only differences in the process.

Of course, as any other method to investigate insight problem solving, FoW ratings have limitations. The criticism has been raised that subjective ratings may be disconnected from actual solution processes (Weisberg, 1992). The present data do not support this idea. Subjective FoW ratings were clearly predictive of solution accuracy, with later solvers (correct solutions) giving consistently higher ratings than later non-solvers (no solutions). This is in accordance with a recent study by Kizilirmak et al. (2018) who found very similar, flat FoW rating curves for unsolved problems, but also with other findings which showed that subjective Aha! ratings are also predictive of solution correctness (Danek et al., 2014b; Hedne et al., 2016; Salvi et al., 2016; Threadgold et al., 2018; Webb et al., 2016; see Danek & Salvi, 2018, for a discussion). Further, the Kizilirmak study (2018) could show a connection between these two subjective measures, FoW and Aha! ratings, by demonstrating that solutions accompanied by an Aha! experience were characterised by a steeper, more sudden increase in FoW ratings than solutions where no Aha! was reported.

The present study contributes to the demystification of insightful problem solving (Bowden et al., 2005) by offering tools to guide problem solvers into and out of an impasse. The ability to influence the problem solving process can serve both educational and research purposes. Gaining insight into difficult problems is part of studying, and teachers should be aware of factors which impede their students' progress, as well as how to guide them to overcome such challenges. Based on the present findings, taking away incorrect solution ideas could be helpful for having students realise that they are in an impasse, or rather that they do not know the answer to a specific question yet. This might make them more open to actively search for new information, instead of continuing to entertain their incorrect ideas due to lack of feedback. A remaining problem that needs to be countered is a possible fixation on the incorrect ideas even after feedback.

Presenting cues led to a clear increase in solution rates, but not to performance at ceiling, which demonstrates that giving helpful but not too revealing cues may help students to realise the solution of difficult problems by themselves. We could also show that cues which match the modality of the presented problem can enhance problem solving performance.

Further, the creation of useful solution cues could also advance the neuroscientific investigation of insight problem solving. According to Luo and Knoblich's (2007) outline, the ability to elicit multiple insight events is crucial for comprehensive hypothesis testing. Displaying solution cues is one possibility to do this and also to better control for when an insight will happen. The ability to manipulate the sequence of the problem-solving process brings researchers a step closer to disentangling how higher cognitive processes are represented in the brain.

We conclude that the FoW rating procedure is a useful way of measuring participants' subjective experience of impasse across time and that FoW ratings were predictive of later task performance, with characteristic differences between solvers and non-solvers. Further, it was possible to manipulate participants' problem solving process, guiding them to and through an impasse by first taking away trivial solution candidates and then giving cues to solution. Taken together, these results provide empirical support for the representational change theory of insight (Ohlsson, 1992) by showing that problem solvers are led into an impasse by depleting trivial solution ideas, and that implementing cues does lead out of the impasse and towards a solution, by helping problem solvers relax their constraints.

Acknowledgments

We thank Franziska Konitzer and Joshua Yudice for writing the code for this experiment.

Disclosure statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research project was funded by a grant to Amory H. Danek from LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative.

ORCID

Judit Pétervári http://orcid.org/0000-0001-5414-3670

References

Abbott, A. (2005). Physics and the public: Science as illusion. *Nature*, *434*(7035), 820–820. doi:10.1038/434820a

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

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. doi:10.1016/j.jml.2007.12.005

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. doi:10.1016/j.tics.2007.05.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. ArXiv:1406.5823 [Stat]. Retrieved from http://arxiv.org/abs/1406.5823

Beeftink, F., van Eerde, W., & Rutte, C. G. (2008). The effect of interruptions and breaks on insight and impasses: Do you need a break right now? *Creativity Research Journal*, *20*(4), 358–364. doi:10.1080/10400410802391314

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

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 634–639. doi:10.3758/BF03195543

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

Chabani, E., & Hommel, B. (2014). Effectiveness of visual and verbal prompts in training visuospatial processing skills in school age children. *Instructional Science*, *42*(6), 995–1012. doi:10.1007/s11251-014-9316-7

Chein, J. M., Weisberg, R. W., Streeter, N. L., & Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Memory & Cognition*, *38*(7), 883–892. doi:10.3758/MC.38.7.883

Connell, L., & Lynott, D. (2011). Modality switching costs emerge in concept creation as well as retrieval. *Cognitive Science*, *35*(4), 763–778. doi:10.1111/j.1551-6709.2010.01168.x

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. doi:10.7771/1932-6246.1129

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

Dahl, D. W., & Moreau, P. (2002). The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, *39*(1), 47–60. doi:10.1509/jmkr.39.1.47.18930

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

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2013). Aha! experiences leave a mark: Facilitated recall of insight solutions. *Psychological Research*, *77*(5), 659–669. doi:10.1007/s00426-012-0454-8

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

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014b). Working wonders? Investigating insight with magic tricks. *Cognition*, *130*(2), 174–185. doi:10.1016/j.cognition.2013.11.003

Danek, A. H., & Salvi, C. (2018) Moment of truth: Why Aha! experiences are correct. *The Journal of Creative Behavior*. Advance online publication. doi:10.1002/jocb.380.

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

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

Davidson, J. E. (1995). The suddenness of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 125–155). Cambridge, MA: MIT Press.
Derbentseva, N. (2007). *The intensity of the insight experience in problem solving: Structural and dynamic properties*. Ontario: University of Waterloo.
Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Springer.
Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5), i–113..doi: 10.1037/h0093599
Fedor, A., Szathmáry, E., & Öllinger, M. (2015). Problem solving stages in the five square problem. *Frontiers in Psychology*, *6*, 1050. doi:10.3389/fpsyg.2015.01050
Fleck, J. I., & Weisberg, R. W. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory & Cognition*, *32*(6), 990–1006. doi:10.3758/bf03196876
Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, *25*(4), 436–463. doi: 10.1080/20445911.2013.779248
Gick, M. L., & Lockhart, R. S. (1995). Cognitive and affective components of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 197–228). Cambridge, MA: MIT Press.
Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving. *Psychological Science*, *14*(5), 462–466. doi:10.1111/1467-9280.02454
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. doi:10.3389/fpsyg.2016.01314
Jausovec, N., & Bakracevic, K. (1995). What can heart rate tell us about the creative process? *Creativity Research Journal*, *8*(1), 11–24. doi:10.1207/s15326934crj0801_2
Katona, G. (1940). *Organizing and memorizing: Studies in the psychology of learning and teaching*. New York, NY: Columbia University Press.
Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 3–13. doi:10.1037/0278-7393.30.1.3
Kizilirmak, J. M., Serger, V., Kehl, J., Öllinger, M., Folta-Schoofs, K., & Richardson-Klavehn, A. (2018). Feelings-of-warmth increase more abruptly for verbal riddles solved with in contrast to without Aha! experience. *Frontiers in Psychology*, *9, 1404*. doi:10.3389/fpsyg.2018.01404
Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534–1555. doi:10.1037/0278-7393.25.6.1534
Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, *29*(7), 1000–1009. doi:10.3758/BF03195762
Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*(1), 71–93. doi:10.1146/annurev-psych-010213-115154
Kuhn, G., Amlani, A. A., & Rensink, R. A. (2008). Towards a science of magic. *Trends in Cognitive Sciences*, *12*(9), 349–354. doi:10.1016/j.tics.2008.05.008
Laukkonen, R. E., & Tangen, J. M. (2017). Can observing a Necker cube make you more insightful? *Consciousness and Cognition*, *48*, 198–211. doi:10.1016/j.concog.2016.11.011
Laukkonen, R. E., & Tangen, J. M. (2018). How to detect insight moments in problem solving experiments. *Frontiers in Psychology*, *9, 282*. doi:10.3389/fpsyg.2018.00282

Luo, J., & Knoblich, G. (2007). Studying insight problem solving with neuroscientific methods. *Methods*, *42*(1), 77–86. doi:10.1016/j.ymeth.2006.12.005

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 176–201. doi:10.1037//0278-7393.27.1.176

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

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review*, *2*(2), 100–110. doi:10.1207/s15327957pspr0202_3

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

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

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

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

Öllinger, M., Jones, G., Faber, A. H., & Knoblich, G. (2013). Cognitive mechanisms of insight: The role of heuristics and representational change in solving the eight-coin problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 931–939. doi:10.1037/a0029194

Öllinger, M., Jones, G., & Knoblich, G. (2014a). Insight and search in Katona's five-square problem. *Experimental Psychology*, *61*(4), 263–272. doi:10.1027/1618-3169/a000245

Öllinger, M., Jones, G., & Knoblich, G. (2014b). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychological Research*, *78*(2), 266–275. doi:10.1007/s00426-013-0494-8

Ormerod, T. C., MacGregor, J. N., & Chronicle, E. P. (2002). Dynamics and constraints in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 791–799. doi:10.1037//0278-7393.28.4.791

Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory & Cognition*, *29*(2), 214–221. doi:10.3758/BF03194915

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

R Development Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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

Sas, C., Luchian, E., & Ball, L. (2010). *Investigating visual analogies for visual insight problems*. Paper presented at the Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design, 90–100. Retrieved from http://dl.acm.org/citation.cfm?id=1854985

Schönauer, M., Brodt, S., Pöhlchen, D., Breßmer, A., Danek, A. H., & Gais, S. (2018). Sleep does not promote solving classical insight problems and magic tricks. *Frontiers in Human Neuroscience*, *12*, 72. doi:10.3389/fnhum.2018.00072

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

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 33–62). Cambridge, MA: MIT Press.

Slepian, M. L., Weisbuch, M., Rutchick, A. M., Newman, L. S., & Ambady, N. (2010). Shedding light on insight: Priming bright ideas. *Journal of Experimental Social Psychology*, *46*(4), 696–700. doi:10.1016/j.jesp.2010.03.009

Smith, S. M. (2003). The constraining effects of initial ideas. In P. Paulus & B. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 15–31). Oxford, UK: Oxford University Press.

Smith, S. M., & Blankenship, S. E. (1989). Incubation effects. *Bulletin of the Psychonomic Society*, *27*(4), 311–314. doi:10.3758/BF03334612

Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *The American Journal of Psychology*, *104*(1), 61–87.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837–845. doi:10.3758/BF03202751

Storm, B. C., & Angello, G. (2010). Overcoming fixation: Creative problem solving and retrieval-induced forgetting. *Psychological Science*, *21*(9), 1263–1265. doi:10.1177/0956797610379864

Thomas, L. E., & Lleras, A. (2009). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin & Review*, *16*(4), 719–723. doi:10.3758/PBR.16.4.719

Threadgold, E., Marsh, J. E., & Ball, L. J. (2018). Normative data for 84 UK English rebus puzzles. *Frontiers in Psychology*, *9, 2513*. doi:10.3389/fpsyg.2018.02513

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*(6972), 352–355. doi:10.1038/nature02223

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

Weisberg, R. W. (1992). Metacognition and insight during problem solving: Comment on Metcalfe. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 426–431. doi:10.1037/0278-7393.18.2.426

WMA's Declaration of Helsinki serves as guide to physicians. (1964). *JAMA*, *189*(13), 33–34. https://doi.org/10.1001/jama.1964.03070130073046

Žauhar, V., Bajšanski, I., & Domijan, D. (2016). Concurrent dynamics of category learning and metacognitive judgments. *Frontiers in Psychology*, *7, 1473*. doi:10.3389/fpsyg.2016.01473