Behav Res
DOI 10.3758/s13428-017-0972-9

Once more with feeling: Normative data for the aha experience in insight and noninsight problems

Margaret E. Webb[1,2] · Daniel R. Little[1] · Simon. J. Cropper[1]

Abstract Despite the presumed ability of insight problems to elicit the subjective feeling of insight, as well as the use of so-called insight problems to investigate this phenomenon for over 100 years, no research has collected normative data regarding the ability of insight problems to actually elicit the feeling of insight in a given individual. The work described in this article provides an overview of both classic and contemporary problems used to examine the construct of insight and presents normative data on the success rate, mean time to solution, and mean rating of aha experience for each problem and task type. We suggest using these data in future work as a reference for selecting problems on the basis of their ability to elicit an aha experience.

Keywords Insight problem solving · Creativity · Aha experience

The feeling of sudden clarity and understanding, often accompanied by a sub-vocal or exuberantly shouted "aha," is known to many as *insight* in problem solving contexts. This feeling of insight (also known as an *aha experience*) has been shown to both improve motivation in problem solving (Liljedahl, 2005) and facilitate recall (Danek, Fraps, von Müller, Grothe, & Öllinger, 2013; Kizilirmak, Gomes da Silva, Imamoglu, & Richardson-Klavehn, 2016). Despite these benefits, finding methods that reliably test insight is a recognised challenge (Bowden, Jung-Beeman, Fleck, & Kounios, 2005). For instance, although the investigation of insight has had a long history (Duncker, 1945; see, e.g., Gilhooly & Murphy, 2005; Jung-Beeman et al., 2004; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Köhler, 1921; Maier, 1931; Metcalfe, 1986a; see also Sternberg & Davidson, 1995), only a handful studies have investigated which specific problems reliably elicit the feeling of insight (Davidson, 1995; Metcalfe, 1986b; Metcalfe & Wiebe, 1987). Furthermore, those studies that have investigated insight problem reliability were predominantly concerned with the subjective experiences (e.g., the feeling of warmth to a solution) leading to solution in insight and noninsight problems. The goal of the present article is to provide a detailed investigation of the strength and reliability of a range of problems used to elicit the cognitive processes and affective components of insight in the individual solving the problem.

Recent research in cognitive neuroscience has demonstrated that problems currently used as insight problems (i.e., compound remote associates, remote associate problems, and anagrams) can elicit insight and noninsight responses (Aziz-Zadeh, Kaplan, & Iacoboni, 2009; Bowden & Beeman, 1998; Kounios et al., 2008; see Kounios & Beeman, 2014, for a review; see also Luo & Knoblich, 2007). However, these studies investigate insight as a categorical response (i.e., participants indicate whether the solution occurred to them through insight or noninsight). Consequently, this method does not reveal the strength of the *aha* experience elicited. A more recent investigation (Danek, Wiley, & Öllinger, 2016) examined the strength of insight elicited by insight problems but only for three classic problems, making the results difficult to generalize. Here we test a wide array of different problems using a continuous measure of insight strength.

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13428-017-0972-9) contains supplementary material, which is available to authorized users.

✉ Margaret E. Webb
mbwebb@student.unimelb.edu.au

[1] University of Melbourne, Parkville, Victoria, Australia

[2] Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria 3056, Australia

Although insight is a subjective experience, there are a number of good reasons to study it. Aside from evidence indicating that the experience of insight is common (Jarman, 2014; Ovington, Saliba, Moran, Goldring, & MacDonald, 2015) and thus of significant general interest, insight has been associated with new and innovative thinking (Feynman, 1999; Poincaré, 1913; Schultz, 1890), facilitated recall (Danek et al., 2013; Kizilirmak, Thuerich, Folta-Schoofs, Schott, & Richardson-Klavehn, 2016), improved learning (Dominowski & Buyer, 2000; Kizilirmak, Gomes da Silva, et al., 2016), and increased motivation (Liljedahl, 2004, 2005). For example, Kizilirmak, Thuerich, et al. (2016) presented participants with a series of compound remote associates (a type of insight problem that has been increasingly used, particularly in the cognitive neuroscience literature). Aha experiences during encoding predicted a significantly higher proportion of solutions to be both recalled and recognised during subsequent testing, presumably due to the deeper encoding afforded by the sudden realisation of the relation between the word problems. To take a second example, Liljedahl (2005) evaluated the impact of the aha experience on motivation for learning mathematics; students who had had an aha experience became less anxious about mathematics and more willing to continue through a problem solving process until they had reached the solution. The investigation of insight delves into the processes underlying these problem-solving techniques, the understanding of which may aid creative problem solving, motivation in learning, and memory.

Q1 Defining insight

Definitions of insight can be approached in three ways: (1) the *process-based* approach, which is concerned with the cognitive processes involved in problem solving; (2) the *task-based* approach, which is concerned with identifying problems that are capable of eliciting insight, with much of this approach being used to determine insight problems that elicit insight processes (e.g., Davidson, 1995; Metcalfe & Wiebe, 1987; Weisberg, 1995b); and (3) *phenomenological* approaches, which are focused more on the feeling of insight (Chronicle, MacGregor, & Ormerod, 2004). Both task and process approaches to insight require an understanding of the *problem space* associated with each problem; that is, the mapping of all possible steps from an unresolved question or issue to the solution. When the steps from one point in the problem space to the next are clear, problem solving is able to progress in steady, incremental, steps. However, in instances when the steps toward solution are not clear, problem solving becomes discontinuous (Weisberg, 1995b); that is, there is a need to wait until further thought about the problem reveals or clarifies the solution process, or until a mental *restructuring* occurs (Ohlsson, 1984; Sandkühler & Bhattacharya, 2011). The term restructuring implies that the way an individual perceives or conceives a problem, and possibly the solution pathway, is fundamentally changed (Weisberg, 1995b). It is this sudden restructuring that is presumed to elicit the phenomenological component of insight (Cushen & Wiley, 2012; Fleck & Weisberg, 2004). In contrast, from a problem-space perspective, a noninsight problem is a problem that does not require restructuring because all problem-solving steps are known from the outset, or at least follow logically from the first step.

Cognitive restructuring is a fundamental aspect of contemporary research on insight (e.g., Ash & Wiley, 2006; Cushen & Wiley, 2012; Sandkühler & Bhattacharya, 2011; Weisberg, 1995a), which focuses on (1) the psychological response leading to and resulting from restructuring of a problem space (Ash & Wiley, 2006); (2) the use of heuristics (Chronicle et al., 2004; Öllinger, Jones, Faber, & Knoblich, 2012); and (3) progress monitoring (in which a problem solver attempts to minimize the gap between the current state of the problem and the goal state; see, e.g., Jones, 2003; MacGregor, Ormerod, & Chronicle, 2001). In process-based approaches, the solution of an insight problem is often presumed to indicate insight, which in turn depends upon the definition of the problem itself.

Task-oriented approaches to defining insight are similarly concerned with designing or identifying those problems that require restructuring for their solutions (i.e., insightful processing). This is often achieved by creating a problem with an initially uncertain or unusual path from problem to solution (i.e., an ill-defined problem space), perhaps by encouraging a faulty initial representation of the problem, through the over-representation of problem constraints (i.e., subjects are encouraged to believe that the problem includes constraints that are not there), infrequent word use, uncommon object use, or suggestive instruction. Insight tasks (insight problems) are then compared to tasks that require incremental solutions (see the supplementary materials for a selection of insight and noninsight problems).

Finally, a phenomenological approach to defining insight focuses on the experience of insight, including the emotional components of that experience (Danek, Fraps, von Müller, Grothe, & Öllinger, 2014a; Shen, Yuan, Liu, & Luo, 2016), and what might elicit or predict those feelings (Topolinski & Reber, 2010). This area of research has grown abruptly in the last decade, with a number of researchers noting the somewhat circular reasoning of terming insight problems as "problems that require insight," and inferring that " insight occurs when insight problems are solved" (Öllinger & Knoblich, 2009, p. 277). To break this circularity, investigators have used self-report to determine whether a given question has elicited an Q2 experience of insight or otherwise (Bowden & Jung-Beeman, 2003a; Danek et al., 2014a; Danek et al., 2016). These self-reports may be gathered either during problem solving (e.g., Metcalfe & Wiebe, 1987) or directly after problem solving (e.g., Bowden & Jung-Beeman, 2003a; Danek et al., 2014a; Kounios et al., 2008). In the present article, we have opted to use the post-problem self-report scales developed by Danek et al. (2014a), which are concerned with the phenomenological components of insight; namely confidence, aha experience, surprise, pleasure, impasse.

One of the most distinctive components of an experience of insight is the *aha experience*. The aha experience has been used as a synonymous term for insight; it is generally described as sudden, accompanied by strong emotional arousal that may be either positive or negative (Danek et al., 2014a; Hill & Kemp, 2016b; Shen et al., 2016), as well as a strong sense of certainty in the reanalysis of the problem. A number of researchers consider the aha experience to be definitive of an insightful solution (Cushen & Wiley, 2011; Gick & Lockhart, 1995; Metcalfe & Wiebe, 1987), or at least the most indicative characteristic of insight problem solving (Danek et al., 2014a; Faber, 2012; Jung-Beeman et al., 2004; Schooler, Ohlsson, & Brooks, 1993).

Much of the cognitive neuroscience literature on insight has focused on validating the procedure developed by Bowden (1997), who solicited trial-by-trial judgments from participants regarding whether a solution was derived through a process of insight or through a process of analysis. Bowden (1997) found that the conscious awareness of insight processes is related to unconscious processing prior to the experience of insight (i.e., when solution words are presented subliminally, solutions are rated by participants as feeling insightful). Subsequent research using this procedure has indicated that the number of solutions that have involved insight varied with distinct brain activations (Jung-Beeman et al., 2004; Kounios et al., 2006; Subramaniam, Kounios, Parrish, & Jung-Beeman, 2009), with specific areas associated with distinct stages of preparation for problem solving. However, these trial-by-trial procedures for measuring insight have consistently used binary or categorical classifications of response [e.g., "Was this problem solved: (1) with insight, (2) not with insight, (3) unsure."]; consequently, investigating the potential strength of the insight response has been curtailed by investigating differences in physiological measures (Hill & Kemp, 2016a). Although these (hopefully) should correlate, there is no evidence that this is the case.

Finally, some researchers have considered the aha experience sufficient to define insight (Gick & Lockhart, 1995; Kounios & Beeman, 2009), whereas others dissociate the aha experience from the experience of insight (Danek et al., 2014a; Sandkühler & Bhattacharya, 2011), arguing that insight comprises many components (e.g., surprise, confidence and impasse; Danek et al., 2014a), of which a feeling of aha is only one (Danek et al., 2014a; Danek, Fraps, von Müller, Grothe, & Öllinger, 2014b; Klein & Jarosz, 2011). Yet others consider the aha experience to be a mere epiphenomenon of restructuring the problem space (Ormerod, MacGregor, & Chronicle, 2002; Sandkühler & Bhattacharya, 2011; Weisberg & Alba, 1981). Irrespective of this debate, the aha experience is a strong emotional marker that has been associated with new discoveries (Feynman, 1999; Poincaré, 1913; Schultz, 1890), facilitated recall (Danek et al., 2013), improved learning (Dominowski & Buyer, 2000; Kizilirmak, Gomes da Silva, et al., 2016), and increased motivation (Liljedahl, 2004). As such, it is worthy of study regardless of whether it is necessary and or sufficient as an indicator of an insight experience. In this article, we investigate the validity of a number of commonly used insight and noninsight tasks by testing each problem's ability to elicit insight.

Tasks used to elicit insight and their controls

Insight problems are designed to elicit a feeling of impasse, or being stuck, by creating a problem with an uncertain or unusual path from problem to solution (a so-called ill-defined problem space). For example:

> *A man is escaping from a 60-m tower. He has a length of rope that is 30 m long. He cuts the rope in half, ties it together again, and uses it to escape. How does he do this?*

The answer may or may not be immediately clear; however, the solution becomes obvious if one thinks about cutting the rope along its length rather than its width.[1] It is this sudden clarity of solution and feeling of aha that is used as an indication of insight processes. However, the initial misinterpretation and consequent misrepresentation of problem space varies across observers, as problem solvers are able to solve these problems using both logical deductions and mental leaps toward a solution (Weisberg, 2014). Q3

In contrast, noninsight problems are designed to be solvable in a simple and incremental process, with a clear path through the problem space from the initial problem to the solution. A classic example of noninsight problems are logic-based questions, though there are also many examples using fluid intelligence tasks (such as Raven's Advanced Progressive Matrices; Raven, 2000):

[1] The rope/prisoner problem highlights some issues with scoring insight problems. In this case, the more knowledgeable you are about rope, the more difficult this problem becomes. Since most common rope is made of three strands twisted together, the rope would be very difficult to cut in half lengthways. In this case, higher crystallized knowledge would be detrimental, highlighting the fact that although insight problems were developed in order to be answerable with the same level of prior knowledge, differences in prior knowledge will affect the ease with which one can generate a solution and, presumably, the experience of insight. For instance, a person with no prior experience with rope might experience insight at the solution, whereas a rope expert might or might not experience insight at realizing the problem required ignoring quite common properties of the rope (e.g., its twisted strands).

Bob's father is three times as old as Bob. They were both born in October. Four years ago, he was four times older. How old are Bob and his father?

The solution (Bob is 12; his father is 36) requires basic arithmetic (3 × 12 = 36; 36 – 4 = 32; 4 × 8 = 32); however, although this question arguably requires simply stepping through the arithmetic, it does require a problem solver to remember their basic maths, and not to get caught by multiplying the three and four to get a 12-year-old father, which is actually a frequent response. Thus, the sudden memory of *how* to solve the problem may result in a feeling of insight. The tendency for problem solvers to solve insight problems using both insightful and analytic methods and feelings was made particularly clear in recent research by Danek et al. (2016), who tested three classic insight problems and found problem solvers would solve these problems both with and without insight affect.

Types of insight and noninsight problems

So far we have discussed predominantly "classic problems" (so dubbed by Cunningham, MacGregor, Gibb, & Haar, 2009); however, although these problem types were initially the most frequently used, they have been superseded in recent years by other problem tasks, such as compound remote associates, anagrams, matchstick arithmetic, and rebus puzzles (Table 1 provides an outline of the problem types, along with links to the studies introducing these into the literature or to normative studies, where available). The majority of research into the ability of insight problems to actually elicit insight has been conducted on compound remote associates (see, e.g., Jung-Beeman et al., 2004; Kounios et al., 2006; Salvi, Bricolo, Bowden, Kounios, & Beeman, 2016; Sandkühler & Bhattacharya, 2011; Wegbreit, Suzuki, Grabowecky, Kounios, & Beeman, 2012), but many of the theories around insight processes arise from research in classic problems (see Sternberg & Davidson, 1995, for a comprehensive review of this literature). We next review classic insight problems and more contemporary insight problems such as the aforementioned, compound remote associates (but also several other more contemporary problem types).

Classic problems

The example above (i.e., the rope problem) is an example of a classic insight problem. These are often riddle-type vignettes, sometimes accompanied by images to create a spatial problem (see Supplementary Materials for list of problems and solutions). Classic insight problems are typically described as impossible to solve without restructuring (Ash & Wiley, 2006; Gilhooly & Murphy, 2005; Weisberg, 1995b). That is, developing a mental representation of the problem that considers the relations between the elements of the problem in a way other than as presented. Weisberg (1995a, b) developed a taxonomy of insight and noninsight problems, based on the degree of restructuring required, and whether or not a problem was discontinuous (whether a problem solver needs to change direction/start again in order to proceed). This taxonomy outlines "pure" noninsight problems for which no restructuring is required, "pure" insight problems, which are both discontinuous and require restructuring, and hybrid problems, which are discontinuous and may require restructuring on a subject-to-subject basis. Gilhooly and Murphy (2005) compared performance on 24 presumed insight and ten presumed noninsight problems in a cluster analysis and found clusters that were congruent with Weisberg's (1995b) taxonomy, including hybrid problems. Q22

The other example presented above (i.e., *Bob's father*) is of a classic noninsight problem, and a large literature has been concerned with testing the procedural differences between classic insight and noninsight problems (e.g., Gilhooly & Murphy, 2005; Metcalfe & Wiebe, 1987; Weisberg, 1995b). However, there are instances in which problems classified as "noninsight" have been solved with insight-like feelings or patterns of solution (e.g., Davidson, 1995; Webb, Little, & Cropper, 2016b). For example, Davidson (1995) noted 12–13% of noninsight problems were solved with the same FOW (feeling of warmth) ratings as insight problems. Webb, Little, and Cropper (2016a, b) investigated a subset of classic insight and noninsight problems, and found that, as with compound remote associates, noninsight problems may also be solved with feelings of insight.

Contemporary problems

In this context, we are distinguishing between classic and contemporary problems in the following fashion: Classic problems are riddles and puzzles drawn from literature and discussed in literature before or during 1995. Classic problems predominately have a vignette component (either as the entirety of the problem, or accompanying a spatial puzzle), and require at least 3 min on average to solve. In contrast, contemporary problems are those that have been developed or discussed predominately after 1995. These include problems such as compound remote associates (Bowden & Jung-Beeman, 2003b), anagrams (Kounios et al., 2008), and rebus puzzles (MacGregor & Cunningham, 2008). We differentiate these from classic problems as, though these problems have been used in the cognitive literature prior to 1995, they have only been applied to the study of insight more recently (see Bowden et al., 2005, for a discussion on this topic).

Compound remote associates and remote associate tasks Both compound remote associates (Bowden & Jung-Beeman, 2003b) and remote associate tasks (Mednick, 1962) are short

Q4 **Table 1** Types of insight and noninsight problems, examples, and directions for further reading

Type	Description	Category	Example	Example Papers
Classic Problems				
Insight	Riddle-type verbal/ arithmetical vignette and figure	Verbal Spatial Mathematical	Marrying man Nine dot problem Egg timer	(Davidson, 1995; Dow & Mayer, 2004; Gilhooly & Fioratou, 2009; Metcalfe, 1986a; Metcalfe & Wiebe, 1987; Weisberg, 1995b) Q5
Noninsight	Logic-type verbal/ arithmetical vignette	Verbal Spatial Mathematical	Dinner problem Cards Water jug	
Contemporary Problems				
Ambiguous images / visual insight task	Visual stimulus that can be interpreted in two ways	Visual	Necker cube	(Laukkonen & Tangen, 2017; Riquelme, 2002; Wiseman, Watt, Gilhooly, & Georgiou, 2011) Q6 Q7
Anagrams	A scrambled word to be unscrambled into a meaningful word	Verbal	uctos = scout	(Bowden, 1997; Metcalfe, 1986b)
Analogies	A verbal relations task in which two words with a relationship are presented, followed by a third word. The task is to find the fourth word that is related to the third.	Verbal	MONTH is to YEAR as HOUR is to ____	(Ansburg, 2000; Qiu, Li, Yang, et al., 2008) Q9/Q8
Arithmetic	Multistep arithmetic problem	Mathematical	Supplementary materials	(Ash & Wiley, 2008; Leikin, Waisman, & Leikin, 2016; Liljedahl, 2004; Topolinski & Reber, 2010) Q10 Q11
Verbal arithmetic	Multistep arithmetic problems presented in a written form.	Mathematical	At Lucky, butter costs 65 cents per stick. Butter at Vons costs 2 cents more per stick than butter at Lucky. If you need to buy 4 sticks of butter, how much will you pay at Vons?	(Thevenot & Oakhill, 2005, 2006, 2008) Q12
Chinese riddles	A phrase, riddle, or poem is presented, and the answer is a single character	Verbal	有口难言	(Qiu et al., 2006; Qiu, Li, Jou, Wu, & Zhang, 2008) Q13
Compound remote associates	Three words that combine together with a single fourth word to create compound words.	Verbal	age/mile/sand	(Bowden & Jung-Beeman, 2003b)* Q14 (Salvi, Costantini, Bricolo, Perugini, & Beeman, 2016)* Q15
Droodles	Nonsensical pictures difficult to understand without being given the theme or verbal clue	Visual		(Nishimoto, Ueda, Miyawaki, Une, & Takahashi, 2010)* Q16
Gear rotation pathways	Problem solving task in which participants predict the turning direction of a final gear in a pathway, given the turning direction of the first gear	Mathematical / Spatial	Supplementary materials	(Stephen, Boncoddo, Magnuson, & Dixon, 2009) Q17
Implicit learning task	A learning task in which participants are given some explicit rules, but during the course of the task will incidentally learn other rules, which are likely to become explicit to participants, but are not explained by the experimenter.	Mathematical	Number reduction task	(Haider & Rose, 2007; Lang et al., 2006) Q18
Magic tricks	Sleight of hand tricks created by magicians.	Visual/spatial		(Danek et al., 2014a)
			III = III + III	(Knoblich et al., 1999)

Table 1 (continued)

Type	Description	Category	Example	Example Papers
Matchstick arithmetic	Matchsticks are presented (either in figure or physical form) as an incorrect equation of Roman numerals. The task is to solve by moving one matchstick.	Spatial/ mathematical		
Mooney images	Photographs which have been manipulated to contain only black and white contrasts (no shades of grey). The task is to decipher what is being presented in the image.	Visual	Supplementary materials	(Kizilirmak, Gomes da Silva, et al., 2016)
Remote associate task	Three words that are can be linked to a single fourth word	Verbal	lick/mine/shaker	(Mednick, 1962)
Raven's Advanced Progressive Matrices	A fluid reasoning test in which a set of patterns are presented to the participant. The task is to complete the pattern. This is used as a noninsight problem.	Spatial	Supplementary materials	(Arthur & Day, 1994; Gilhooly, Fioratou, & Henretty, 2010; Raven, 2000*)
Rebus puzzle	Words and visual cues combined to represent a familiar phrase	Verbal/Spatial	SOMething	(MacGregor & Cunningham, 2008)*
Scientific problem	A short vignette is set before the participant, or a historical problem dealt with by scientists, accompanied (experimentally varied) with the prototype information that inspired the solution		**Situation**: When making body armour, the first material thought of was steel thread since steel thread is strong and tough. However, steel body armour is too heavy, and unsuited for combat. People need a type of body armour that is both light and highly strong. **Problem**: How would you make body armour that is both light and highly strong? **Prototype**: Spider silk is a type of bio-fibre, and is extremely light, yet has a high degree of strength, equivalent to 5 times the same volume of steel thread.	(Yang et al., 2016) Q19
Sentence completion task	An ambiguous sentence is presented, with the task for the participant to find the word that makes sense of the sentence.	Verbal	Fortunately, there was a haystack, for the cloth ripped	(Auble, Franks, Soraci, Soraci, & Soraci, 1979; Luo, Niki, & Phillips, 2004) Q20 /Q21

*These studies present normative data.

verbal problems: Three words are presented to a participant, combinable with a single fourth word. In the case of compound remote associates, the fourth word can combine with the three to create three *compound* words (e.g., *tooth, potato,* and *heart* combine with *sweet*). In the case of the remote associate tasks, the fourth word does not need to create compound words, but is simply related to the three problem words (e.g., *lick, sprinkle,* and *mine* with *salt*). These words have gained prominence in the insight literature because they are relatively short problems, can be easily administered, and have many easily created variations. Bowden and Jung-Beeman (2003b) conducted a normative study on 144 compound remote associates providing response times and solution rates. Concurrent research (Bowden & Jung-Beeman, 2003a) provided evidence to validate the ability of compound remote associates to elicit insight affect and processes; however, information regarding the probability of experiencing insight was not provided.

Anagrams Anagrams are words that have been scrambled and presented to a participant for solution (e.g., tpoil = pilot). Metcalfe (1986b) used these in her research investigating insight-based and analytic-based (i.e., not involving insight) solutions. However, despite subjects indicating that these problems were predominately solved with a feeling-of-warmth rating similar to that experienced in insight problems (i.e., feeling-of-warmth ratings suddenly leap from far to near in insight problems, whereas they incrementally increase in noninsight problems), researchers have presented arguments against the classification of anagrams as insight problems. For instance, Weisberg (1995b) argued that anagrams were not insight problems because they do not require restructuring but rather are a simple vocabulary search task.

Nevertheless, a number of studies have used anagrams for their ability to elicit insight (e.g., Aziz-Zadeh et al., 2009; Bowden, 1997; Jacobsen, 2016; Kounios et al., 2008; Novick & Sherman, 2003). Although different studies have Q23 provided conflicting information regarding the solvability of anagrams (e.g., Novick & Sherman, 2003), no normative data have been collected for the degree of insight processes or affect elicited by different anagrams.

Raven's Advanced Progressive Matrices The logic pattern-completion puzzles Raven (1985) developed in order to assess fluid reasoning and problem solving abilities (Little, Lewandowsky, & Craig, 2014) have been increasingly used as noninsight problems (e.g., Gilhooly, Fioratou, & Henretty, 2010; Paulewicz, Chuderski, & Nęcka, 2007). Each task comprises a 3 × 3 figure matrix organised according to latent rules, with the task being to deduce the latent rule and select one answer from eight possible answers to complete the pattern. Investigations by Gilhooly and Murphy (2005) indicate that performance on Raven's Advanced Progressive Matrices (Raven's) tasks form clusters with classic noninsight problems, yet the literature consistently demonstrates a positive relationship between Raven's and both classic insight problems (Lin, Hsu, Chen, & Wang, 2012; Nęcka, Żak, & Gruszka, 2016; Paulewicz et al., 2007) remote associate tasks (Chermahini, Hickendorff, & Hommel, 2012; Paulewicz et al., 2007). As yet, there have been no investigations into the ability of these tasks to elicit insight or otherwise. Thus, we will investigate their tendency to elicit insight or otherwise in the present study.

Rebus puzzles MacGregor and Cunningham (2008) proposed rebus puzzles as insight problems, obtaining a measure of self-reported insight affect, and comparing performance on rebus puzzles to the remote associate tasks. A rebus puzzle combines words and visual cues to represent a familiar phrase (e.g., SOMething = "the start of something big"). Participants' base ratings of insight were higher in response to rebus puzzles and remote associate tasks as compared to an analogies task (e.g., "*sheep* is to *lamb* as *cow* is to . . ." = calf). These results were interpreted as evidence that rebus puzzles could be considered insight problems. However, MacGregor and Cunningham did not obtain individual insight rating data for their problem sets. Salvi et al. (2016) also used a set of Italian rebus puzzles and found that solutions solved with insight were judged to be correct more often than solutions solved analytically. Salvi et al. replicated these findings for anagrams and compound remote associates, but did not provide data regarding ratings of insight.

Matchstick arithmetic Matchstick arithmetic problems were proposed as insight problems by Knoblich et al. (1999) to investigate the role of chunked information and restructuring. In a matchstick arithmetic task, an incorrect equation is presented to a participant with matchsticks creating both numbers (Roman numerals) and mathematical symbols. The task is to make the equation correct by moving one matchstick (e.g., IV = III – I; answer, IV – III = I). In their experiment, Knoblich et al. tested the degree of restructuring required by each type of matchstick arithmetic; however, they did not investigate the phenomenology of insight. Recent investigations of the ability of these tasks to elicit insight affect have provided mixed results (Danek et al., 2016; Derbentseva, 2007).

Magic tricks A novel method used by Danek et al. (2014b) was to investigate insight using magic tricks. In conjunction with a magician, the researchers developed and recorded 40 short tricks, with only one effect and one method, which were scored according to the degree of insight-related affect (i.e., surprise, aha, impasse, confidence, and pleasure) experienced when watching the trick. Although the magic tricks may or may not conform to standard definitions of insight problem (i.e., restructuring), they evidently elicited insight. Since we chose to investigate the most frequently occurring tasks in the literature, we did not investigate magic tricks or rebus puzzles.

Aim of the present work

A number of the studies discussed above contain normative data for the solution rate and response time of a variety of different problem types; however, there are currently no normative data on the *strength* and *frequency* of insight affect elicited by these tasks. The ability of any of the above problems to elicit insight is not in dispute; evidence indicates that many problems can elicit insight for many persons, depending on an individual's focus and reason for problem solving (Klein & Jarosz, 2011; Ovington et al., 2015). It is the *strength* of insight that is elicited across a range of problems that we aim to investigate in this article, as well as the reliability of a subset of problems to elicit insight.

General method

Across four studies, a total of 544 University of Melbourne students (452 female, 92 male; age range = 16–58, mean =

20.34) completed insight and noninsight problem-solving tasks coupled with various additional measures. The primary study was conducted with 101 University of Melbourne students (72 female, 29 male; age range = 17–58, mean = 23.38), who completed the study for payment of $40. Before beginning the study, participants were provided with consent forms detailing the proposed study. We advertised for participants with English as a first language, as a number of problems required high English proficiency and we have previously shown this to be important (Webb, Little, Cropper, & Roze, 2017).

Materials

Classic insight and noninsight problems To generate a dataset of classic problems, we conducted a systematic search of the literature, and noted which problems were most frequently used (see the supplementary materials for search terms and selection criteria, as well as the table detailing which problems were used most frequently).

Problems were categorized as insight or noninsight problems on the basis of published categorizations and taxonomies. There were some contradictions in the usage of particular problems (e.g., trace problems have been used as both insight and noninsight problems). In these instances, we classified each problem according to the cluster analysis performed by Gilhooly and Murphy (2005).

We selected the top 25 most frequently used insight and noninsight problem. Accuracy and RT were recorded. We provide normative data for the solution of these problems in the Appendix.

Raven's Advanced Progressive Matrices Participants completed the truncated Raven's Advanced Progressive Matrices (adapted according to the method of Arthur & Day, 1994), which contains 12 test problems. These 12 problems were randomly interleaved with classic insight and noninsight problems. Accuracy and reaction time were recorded, with normative data for the solution of these problems in the Appendix.

Compound remote associates We presented participants with 34 problems, pseudo-randomly drawn from each quantile in Bowden and Jung-Beeman's (2003b) dataset, ensuring that the solutions would vary in difficulty and time necessary for solution. Participants had 30 s to generate the fourth word.

Anagrams We drew 34 five-letter anagrams from Novick and Sherman (2003). Each anagram was solvable within one-, two-, or three-letter moves for the solution, with two-letter moves being most common.

Procedure

Each participant was individually tested in four sessions. Problems were presented online using Qualtrics (Qualtrics, 2016) to present problems and record reaction times (for more detail on the resolution of reaction time measures in Qualtrics, see Barnhoorn, Haasnoot, Bocanegra, & van Steenbergen, 2014). The problem-solving sets were counterbalanced across participants. No solutions were given.

Problem-solving sets There were two problem-solving sets: classic and contemporary problem solving, respectively. The classic "insight" and "incremental" (noninsight) problems were randomly interleaved within a set. Participants were given no information about whether the problem to be solved was classified as "insight" or "noninsight" but were given 210 s to work through the problem. In the contemporary problem set, compound remote associate and anagram components were counterbalanced. Five practice trials preceded each set. Participants were given 30 s to solve each contemporary problem.

Participants were given information on aha experiences to respond in their ratings to each problem. A vignette describing aha experiences (drawn from Danek, Fraps, von Müller, Grothe, & Öllinger, 2014a, b; see the supplementary materials for the vignette) was presented at the beginning of the experiment. After each problem solving task, participants were presented with the scales drawn from (Danek et al., 2014a). We chose to use these scales as they individuate components of insight from one another, and as a visual analogue they require minimal processing. Participants were asked to rate: (1) the *confidence* that the given response was correct (*very unsure* to *very sure*), (2) the *strength* of the aha experience (*very weak* to *very strong*), (3) the *pleasantness* of the insight experience (*very unpleasant* to *very pleasant*), (4) the *surprising* nature of the insight experience (*not surprising* at all to *very surprising*), and (5) the feeling of *impasse* before the insight experience (*no impasse* at all to *very stuck*). Participants responded by moving a slider (preset at 50) along a scale of 0–100.

Data analysis

Analyses were conducted using JASP (Love et al., 2015) and R. Differences in the aha ratings across problem types were investigated using a series of one-way ANOVAs, whereas the correlation plots were created using the R package corrplot (Wei & Simko, 2016).

Results

Problems were scored as either correct or incorrect and averaged across category (insight, noninsight, compound remote associates, anagrams), as were the ratings of insight-related affect (see the supplementary materials). Descriptive statistics for performance accuracy and the ratings of insight-related affect are displayed in Table 2.

We calculated the percentage of participants solving each problem, as well as the mean time to solution, in seconds. We also calculated the mean ratings of insight for each problem, and then further investigated the mean ratings of aha experience by response accuracy. These data are presented in the Appendix in descending order according to mean strength of insight elicited in correct responses.

Table 2 Descriptive statistics of accuracy and insight related affect across problem types

Problem Type	Mean	*SD*
Insight		
Accuracy	0.30	0.16
Aha	42.04	16.44
Confidence	51.92	15.28
Impasse	57.56	14.29
Pleasure	48.94	14.48
Surprise	40.82	17.87
Noninsight		
Accuracy	0.57	.18
Aha	50.45	17.18
Confidence	67.73	11.95
Impasse	47.19	14.86
Pleasure	58.71	12.84
Surprise	37.32	16.89
Raven's		
Accuracy	0.57	.20
Aha	54.50	18.93
Confidence	74.48	13.47
Impasse	42.27	16.43
Pleasure	62.66	14.49
Surprise	37.12	18.26
Compound Remote Associates		
Accuracy	0.34	0.18
Aha	39.58	17.07
Confidence	45.36	15.21
Impasse	65.17	16.53
Pleasure	43.63	14.64
Surprise	39.13	20.47
Anagrams		
Accuracy	0.78	0.16
Aha	64.17	18.64
Confidence	79.07	13.12
Impasse	38.13	21.01
Pleasure	66.93	15.82
Surprise	34.65	20.14

Relationships between problem types

We examined the relationships between problems used as insight problems (classic insight problems, anagrams, and compound remote associates), and problems used as noninsight problems (classic noninsight problems and Raven's Advanced Progressive Matrices) in terms of both accuracy and the strength of the aha experience.

The correlations between problem types on ratings of aha experience indicated moderate to strong positive relationships across problem types, as can be seen above the diagonal in Fig. 1 (note that all relationships are above a Pearson *r* value of .4 and significant at $p < .001$; below the diagonal are the correlations for accuracy). This indicates that individual differences may underlie the tendency to report a problem to be solved with insight across both insight and noninsight problem types, as has been noted through the use of compound remote associates and anagrams in the cognitive neuroscience literature (Bowden et al., 2005; Kounios & Beeman, 2014).

Performance accuracy The pattern of relationships across problem types in terms of accuracy indicates significant positive relationships between classic insight problems and all other problem types (see the lower half of Fig. 1; also see the supplementary materials for correlation statistics), as well as significant moderate positive relationships between solution accuracy on anagrams and compound remote associates

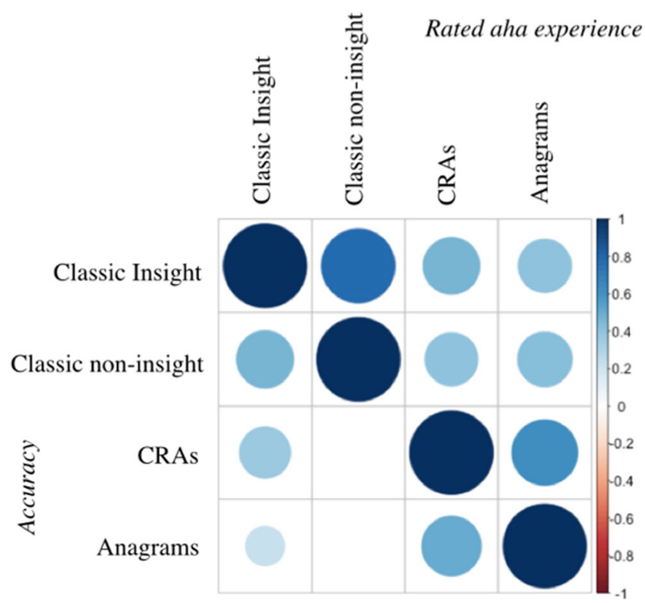

Fig. 1 Correlation plots between accuracy and aha across problem types. The size of each circle and its saturation of color show the strength of the correlation; the color shows the direction of the relationship, with positive being blue. The upper half of the correlation plot details aha results, and the lower half details accuracy. Nonsignificant correlations have been removed (see the supplementary materials for the correlation statistics). The correlation plot was created using the R package corrplot (Wei & Simko, 2016)

[$r(99) = .51$, $p < .001$] and between noninsight problems and both compound remote associates [$r(99) = .25$, $p = .01$]. However, accuracy on noninsight problems was not correlated with anagrams [$r(99) = .18$, $p = .07$]. Furthermore, despite significant positive relationships between Raven's and both insight [$r(99) = .39$, $p < .001$] and noninsight [$r(99) = .56$, $p < .001$] problems, there were no significant relationships between Raven's and either anagrams [$r(99) = -.09$, $p = .39$], or compound remote associates [$r(99) = .06$, $p = .51$]. This may reflect the necessity of an extensive vocabulary for the solution of both compound remote associates and anagrams, whereas Raven's is a nonlexical solution. It also reflects some of the complications of using these problem types interchangeably, as was noted by Ball and Stevens (2009).

Differences between problem types for accuracy and insight

We were also interested in whether particular problem types (e.g., classic insight problems) would elicit higher ratings of insight experience, particularly, ratings of the aha experience. If all problems considered to be insight problems can be used interchangeably, we would expect a significant difference in aha ratings for problems considered to be insight problems (i.e., classic insight problems, compound remote associates, anagrams) as compared to problems considered to be noninsight problems (i.e., classic noninsight problems, Raven's), and no difference between problem types within insight or noninsight categories. A repeated measures analysis of variance on ratings of aha experience across problem types (see Fig. 2) indicated a significant difference between problem types on aha ratings: $F(4, 400) = 65.85$, $p < .001$, $\eta^2 = .40$. Post-hoc comparisons showed no significant difference between insight problems and compound remote associates in aha ratings. This implies that classic insight problems and compound remote associates elicit, on average, ratings of insight that are not significantly different from each other, which is reassuring for a literature that is moving from the use of classic insight problems to compound remote associates.

Similarly, no significant differences emerged between noninsight problems and Raven's in aha ratings, which may indicate that Raven's is a valid measure of noninsight problem solving; however, noninsight problems resulted in significantly higher ratings of aha experience than both insight problems ($p < .001$, mean difference = 6.97, Cohen's $d = 0.62$) and compound remote associates ($p < .001$, mean difference = 10.02, Cohen's $d = 0.55$). (Similarly, Raven's resulted in significantly higher ratings of aha than either insight problems—$p < .001$, mean difference = 11.17—or compound remote associates—$p < .001$, mean difference = 14.22.) These results extend the findings of Danek et al. (2016), who noted that classic insight problems could be solved without insight, with the finding that classic noninsight problems can be solved with *strong* feelings of insight.

Finally, anagrams elicited significantly higher ratings of aha experience than did all other problem types (anagrams to classic insight: $p < .001$, mean difference = 21.55, Cohen's $d = 1.11$; anagrams to compound remote associates: $p < .001$, mean difference = 24.59, Cohen's $d = 1.57$; anagrams to noninsight: $p < .001$, mean difference = 14.58, Cohen's $d = 0.77$; anagrams to Raven's: $p < .001$, mean difference = 10.37, Cohen's $d = 0.48$).

Accuracy Given the process-oriented approach of interpreting the correct solution of an insight problem as indicative of insight, we performed the same repeated measures ANOVA across problem types for solution accuracy (see Fig. 2b). We found a significant difference in accuracy across problem types, $F(4, 400) = 222.40$, $p < .001$, $\eta^2 = .68$, with participants being significantly more accurate at solving anagrams than at solving all other problem types (anagrams to classic insight: $p < .001$, mean difference = .47, Cohen's $d = 2.30$; anagrams to compound remote associates: $p < .001$, mean difference = .44, Cohen's $d = 2.66$; anagrams to noninsight: $p < .001$, mean difference = .17, Cohen's $d = 0.74$; anagrams to Raven's: $p = .004$, mean difference = .07, Cohen's $d = 0.27$). Participants solved significantly more Raven's problems than noninsight ($p < .001$, mean difference = .09, Cohen's $d = 0.99$), insight (p

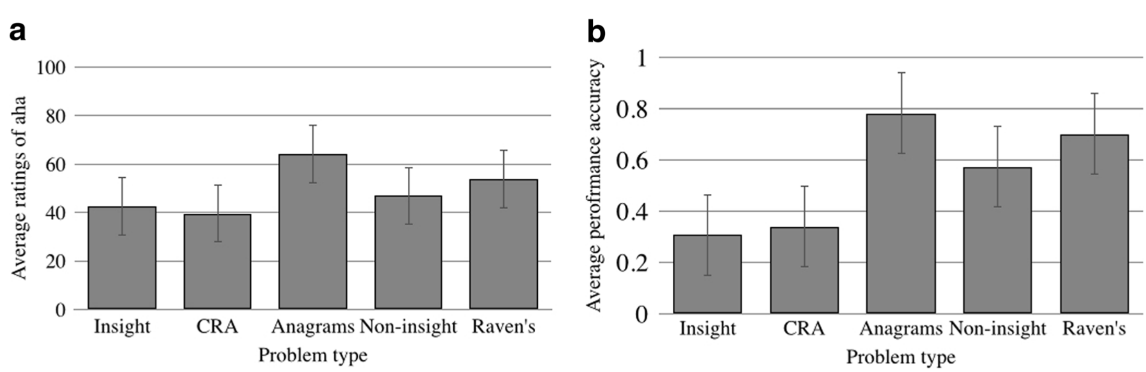

Fig. 2 Mean (**a**) aha ratings and (**b**) performance accuracy across problem types. (Error bars show standard deviations)

< .001, mean difference = .40, Cohen's *d* = 1.92), or compound remote associates (*p* < .001, mean difference = .37, Cohen's *d* = 1.44) problems. They also solved more noninsight problems than either insight problems (*p* < .001, mean difference = .31, Cohen's *d* = 1.81) or compound remote associates (*p* < .001, mean difference = .27, Cohen's *d* = 1.27). We observed no significant difference between insight problems and compound remote associates in accuracy (*p* = .61, mean difference = .03). The results of accuracy reflect the results of ratings of insight, and suggest a relationship between accuracy and aha. The correlations between accuracy and aha ratings (see the supplementary materials, Fig. 1, for correlation plots) indicate a significant relationship between ratings of aha and solution accuracy for presumed insight problems [classic insight: *r*(99) = .27, *p* = .006; compound remote associates: *r*(99) = .26, *p* = .008; anagrams: *r*(99) = .23, *p* = .02], but no relationship for noninsight problems [*r*(99) = – .10, *p* = .31].

Ratings of aha experience conditional on performance accuracy

Given the similarity in the patterns across problems of both aha ratings and accuracy, we performed a series of analyses on aha ratings conditional on whether the problem was correctly solved (see Fig. 3). Looking at the aha ratings across problems when the solution was correct revealed a significant effect of problem type: $F(3, 69) = 29.56$, $p < .001$, $\eta^2 = .56$. Bonferroni post-hoc tests indicated that anagrams elicited the highest ratings of insight relative to other problem types, with significantly higher ratings than classic insight problems (*p* < .001, mean difference = 16.08, Cohen's *d* = 1.11) or classic noninsight problems (*p* < .001, mean difference = 19.69, Cohen's *d* = 2.35) (anagrams were not significantly different from compound remote associates when analyzing aha ratings conditional on correct solutions: *p* = 1, mean difference = 3.30). Compound remote associates returned significantly higher self-reports of aha experience than did either insight problems (*p* < .001, mean difference = 12.78, Cohen's *d* = 0.79) or noninsight problems (*p* < .001, mean difference = 16.39, Cohen's *d* = 2.06).

There was no difference between insight and noninsight problems on ratings of aha for correctly solved problems (*p* = .91, mean difference = 3.61). This suggests that the original finding of significantly higher rating of aha experience for noninsight problems may have resulted from consistently higher ratings of aha for both correct and incorrectly solved noninsight problems, whereas for insight problems, ratings of insight were high only for correctly solved problems. (We found no significant difference between Raven's and noninsight problems: *p* = .19, mean difference = 4.94.)

Across all problem types, a significant difference in aha ratings was apparent for incorrectly solved problems: $F(3, 69) = 11.68$, $p < .001$, $\eta^2 = .33$. Post-hoc comparisons indicated that this significance was driven largely by high ratings of aha experience for incorrectly solved noninsight problems, and low ratings of aha for incorrectly solved compound remote associates (*p* < .001, mean difference = 20.35, Cohen's *d* = 1.66). There was, for instance, no significant difference between the aha ratings for incorrectly solved insight and noninsight problems (*p* = .15, mean difference = 8.02, Cohen's *d* = 0.45), nor between noninsight and Raven's problems (*p* = .14, mean difference = 6.78, Cohen's *d* = 0.62). Ratings of aha were also significantly higher for incorrectly solved noninsight problems than for incorrect anagrams (*p* = .007, mean difference = 11.84, Cohen's *d* = 0.66). Ratings of aha in incorrectly solved compound remote associate problems were also significantly lower than for classic insight problems (*p* = .005, mean difference = 12.33, Cohen's *d* = 0.86).

Summary

We investigated aha ratings across a number of problem types, investigating the relationship between aha and accuracy through correlational analysis and analyses of variance. We found that, when investigating the patterns of differences on

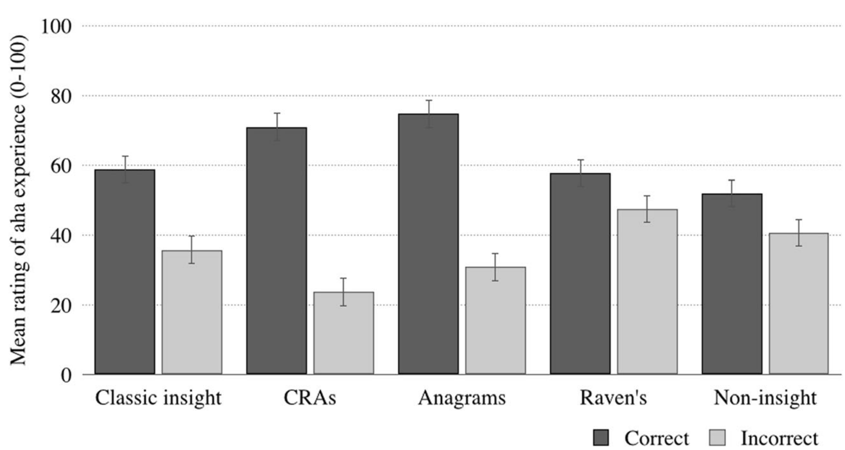

Fig. 3 Mean ratings of aha experience across problem types as a function of accuracy. (Error bars show standard deviations)

average for aha ratings and accuracy individually, the patterns were similar for both accuracy and aha ratings. However, investigating aha ratings across problem types conditional on accuracy provided a different pattern of results.

Overall, anagrams were solved with the highest accuracy and highest ratings of aha. Although compound remote associates also elicited high ratings of aha, their low solution rates mean they are dependent on measuring both accurate solutions and ratings of aha experience.

Interestingly, classic insight problems and noninsight problems were not significantly different in terms of aha ratings analyzed conditionally on response accuracy. Interestingly, aha ratings were significantly *higher* in noninsight problems when not conditionally analyzed. This is in strong contrast to the use of noninsight problems as a control problem (though we recognize that noninsight problems are possibly effective as problems that, more often than not, do not require restructuring). However, we found a significant relationship between accuracy and aha experience in presumed insight problems (compound remote associates, anagrams, and classic insight problems) but no relationship with classic noninsight problems.

We used the truncated Raven's Advanced Progressive Matrices (Arthur & Day, 1994) as noninsight problems. This enabled us to investigate reports of insight-related affect in the solution of Raven's matrices. There was a significant positive relationship between Raven's and all problem types regarding aha experiences, and no significant differences between ratings of aha in classic noninsight problems and Raven's, despite a significant difference in accuracy.

Reliability of classic insight problems to elicit insight

Reliability of insight: Method

In three additional experiments (Webb et al., 2016b; Webb, Little, Cropper, & Roze, 2017), each with large sample sizes ($N > 100$), we used a subset of the problems that we test here using near identical procedures. This allowed us to investigate the reliability of aha ratings conditional on accuracy across all four experiments. The problem set procedure was identical to the method already outlined in this article, with exceptions to this procedure outlined below. The primary focus of these three experiments were to investigate individual differences in the tendency to report insight, and questionnaires were given to participants to complete as well as the problem solving task, in counterbalanced order.

Study 1

Students from the University of Melbourne (193: 118 female, 75 male; age range = 17–52, mean = 19.639) completed the study for course credit. Nine participants were removed for errors on more than 20% of the tasks.

Materials

"Classic" insight and noninsight problems The following problems were used in all studies:

> *Insight problems:* triangle problem, socks problem, lilies problem, antique coin problem, and egg timer problem
> *Noninsight problems:* cards, water jug, trace, police, and dinner

"Contemporary" insight problems: Compound remote associates We used 20 CRAs drawn from Bowden and Jung-Beeman's (2003b).

Questionnaires A series of individual differences measures were presented in random order. These included the Oxford–Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason & Claridge, 2006), Raven's (1985) Advanced Progressive Matrices, a verbal fluency measure adapted from Lezak (2004), and an adaptation of the alternative-uses task (AUT: Guildford, Christensen, Merrifield, & Wilson, 1978). These measures are reported elsewhere in a follow-up study of the same sample (Webb, Little, Cropper, & Roze, 2017).

Study 2

This data was collected individually online. A further aim of this study was to investigate the effect of feedback on reported insight. We only conduct our analysis on the responses taken before the solution was revealed for each problem. The comparison of aha ratings before and after feedback is reported elsewhere in a follow-up study of the same sample (Webb, Little, Cropper, & Webb, 2017).

We found no significant difference in accuracy or aha ratings between the study completed in lab (Study 1) and the study completed online (Study 2).

Participants A total of 129 undergraduates (88 female, 41 male; age range = 17–45, mean = 19.059) completed the tasks for course credit. Twelve participants were removed for errors in more than 20% of the tasks.

Materials, procedure, and design The materials and procedure were identical to Experiment 1, save that participants were given the solution to the problem after their initial attempt.

Study 3

We expanded the individual difference measures in Experiment 3 to include measures of the big five and magical ideation. However, problems and procedure otherwise remained the same as Experiment 1. The tasks were presented individually online.

Participants Undergraduates from the University of Melbourne (130: 106 female, 24 male; age range = 16–47, mean = 19.60) completed the tasks for course credit. Four participants were removed for errors in more than 20% of the tasks.

Reliability of insight: Results

Interexperiment reliabilities were calculated using Cronbach's alpha for each problem type. Ratings of aha experience for correctly solved insight problems were highly reliable, with a reliability coefficient of α = .95. Aha ratings for correctly solved noninsight problems were moderately reliable, with α = .79. For incorrectly solved problems, insight α = .66, noninsight α = .95. The drop in reliability for aha ratings in incorrectly solved problems is congruent with an accuracy-related pattern of aha experience in insight problem solution; that is, feelings of insight were more reliably elicited in participants who correctly solved insight problems. In contrast, even incorrectly solved noninsight problems had a high reliability in aha ratings (i.e., reliably low aha ratings). As is presented in Table 3, we investigated the average aha ratings conditional on accuracy, as well as the problems that could be dropped and increase the reliability of the ratings of aha experience. Within insight problems, the triangle problem was the least reliable for aha ratings in both correctly and incorrectly solved problems. Within noninsight problems, the police problem was the least reliable in terms of aha ratings in both correctly and incorrectly solved problems.

General discussion

We conducted an extensive investigation into the ratings of insight elicited by problems frequently used as tests of both creativity and insight (classic insight and noninsight problems, compound remote associates and anagrams). We recorded measures of solution time, accuracy and ratings of aha experience. We also recorded insight-related affect (e.g., surprise and confidence in solution). The ratings of self-reported insight experience emphasize both the importance of judging insight versus noninsight processes by the feeling in the solution rather than by task (Bowden & Jung-Beeman, 2003a), as well as using the continuous, strength-based, self-report method used in the present study. The results provide support for the use of compound remote associates, anagrams, and classic insight problems as problems that elicit insight; however, they urge caution for the usage of classic noninsight problems and intelligence tests (e.g., Raven's Advanced Progressive Matrices) as controls for insight problems.

Ratings of aha

The present results offer preliminary normative data for the tendency of classic and contemporary insight problems to elicit insight processes and affect.[2] This is particularly useful given the increasing use of compound remote associates to study insight, as existing normative datasets (e.g., Bowden & Jung-Beeman, 2003b) have so far not provided data regarding the tendency of a particular problem to elicit insight affect, only solution rates and reaction time.

Interestingly, ratings of aha experience for anagrams were highest across all problem types; both in the average of reported aha experience and for the problems with correct solutions only. This challenges the perception and use of anagrams as noninsight problems (e.g., Gilhooly & Murphy, 2005; Öllinger, Jones, & Knoblich, 2008; Weisberg, 1995b). For instance, Weisberg (1995b) was concerned that anagrams are a simple memory search task, rather than requiring productive thinking, and so are not true insight tasks. That same critique applies to compound remote associates (Cranford & Moss, 2012), which demonstrably both elicit insight-related affect (Jung-Beeman et al., 2004; Salvi et al., 2016), as well as distinct neurological processes when solved with versus without insight affect (Jung-Beeman et al., 2004; Subramaniam et al., 2009). Interestingly, when analyzed conditionally on accuracy, we found no difference in aha ratings between correctly solved anagrams and compound remote associates. This is congruent with the work of Salvi et al. The high ratings of insight affect in anagrams and compound remote associates may be a consequence of the short solution time required, and the single-word, unambiguous solutions, which may have made the certainty of correct solutions higher and the sense of aha greater (Bowden et al., 2005). Consistent with this, the vignette of Danek et al. (2014a, b) describes insight as being sudden and having a surety of the correctness. In contrast, classic insight and noninsight problems have more ambiguous problem components and solutions, which require holding more information in mind simultaneously.

Despite noninsight problems being used as a control for insight problems (Ash & Wiley, 2006; DeCaro, Van Stockum, & Wieth, 2016; Fleck, 2008; Murray & Byrne, 2005; Wen, Butler, & Koutstaal, 2013; Wieth & Zacks,

[2] By "insight affect" we mean all of the possible components of the feeling of insight, such as surprise, impasse, confidence, pleasure, and the aha experience. By contrast, "aha experience" is used to distinctly refer to our analyses of the aha experience ratings. We have adopted two terms because the aha experience is arguably only one component of insight (see Topolinski & Reber, 2010, for a discussion on this topic).

Table 3 Averaged aha conditional on correct solution of the problem across four experiments

Problem	*Average aha: Correct solution*					*Average aha: Incorrect solution*				
	Cronbach's α (if dropped)	Study 1	Study 2	Study 3	Study 4	Cronbach's α (if dropped)	Study 1	Study 2	Study 3	Study 4
Insight problems	.95					.66				
Triangle	(.96)	73.87	62.18	72.52	74.07	(.56)	32.10	24.18	14.91	30.23
Socks	(.97)	61.16	47.47	49.85	66.28	(.19)	30.69	29.23	25.72	42.92
Lilies	(.92)	75.83	63.57	67.27	76.36	(.90)	26.28	25.54	35.41	30.44
Antique Coin	(.92)	72.41	59.31	67.33	71.76	(.12)	38.40	31.26	31.11	43.69
Noninsight problems	.79					.95				
Cards	(.73)	47.68	45.11	41.91	48.79	(.93)	33.25	22.29	21.15	37.18
Water Jug	(.52)	61.90	52.23	48.95	66.07	(.94)	31.80	23.00	21.78	31.15
Police	(.88)	41.26	36.33	36.09	54.36	(.98)	37.50	37.27	42.00	45.81
Dinner	(.57)	46.28	48.28	50.54	51.12	(.93)	39.36	29.46	32.78	49.57

2011), no significant differences between classic insight and noninsight problems emerged in ratings of aha experience for correctly solved problems. Furthermore, there were actually higher aha ratings in noninsight than in insight problems when aha ratings were averaged over correct and incorrect responses. This may simply reflect the consistently higher aha ratings for both correct and incorrectly solved noninsight problems, whereas insight problems elicited insight predominately for correctly solved problems. These findings are consistent with the thesis that insight problems might be solved incrementally and noninsight problems might be solved insightfully (Bowden, 1997; Danek et al., 2016; Weisberg, 2014). These results call for the use of self-report in all studies investigating insight affect and insight processes (Bowden & Jung-Beeman, 2003a) until the components underlying the phenomenology are better understood.

We investigated the truncated Raven's Advanced Progressive Matrices (Arthur & Day, 1994) as noninsight problems, examining the tendency for the solution of Raven's Matrices to elicit insight affect. Previous studies have found significant positive relationships between Raven's Matrices and both classic insight problems (Lin et al., 2012; Nęcka et al., 2016; Paulewicz et al., 2007) and the precursors to compound remote associates, the remote associate task (Chermahini et al., 2012; Paulewicz et al., 2007). The relationship between Raven's and insight problem solving has been argued to reflect the necessity of fluid reasoning in insight problem solving accuracy (Paulewicz et al., 2007), and we can extend this to note that accuracy is important for high ratings of insight. We have provided data, moreover, to indicate that the solution of Raven's are able to elicit ratings of insight that are not significantly different from those of classic insight problems, which supports a dual-process view of insight problem solving, in which insight can be considered as a normal process, with special add-ons.

On accuracy and insight problem solving

We found positive relationships between accuracy and ratings of aha in presumed insight problems (classic insight problems, anagrams, and compound remote associates), with substantially higher aha ratings for problems with correct solutions. This finding is consistent with the multi-level modeling conducted by Webb, Little, and Cropper (2016b), which showed that insight related affect (i.e., ratings of aha, confidence, and pleasure) were predictive of solution accuracy. From a processing perspective, this finding supports the idea that the solution of presumed insight problems is designed to appear obvious once the problem space has been restructured. Although this supports the idea that restructuring results in an aha experience (Salvi et al., 2016), it is also commensurate with the idea that aha reflects sudden confidence in an answer that is easily verifiable.

One valuable question raised by the present results (and previous results; see Danek et al., 2014a, b; Webb et al., 2016a, b) is whether there is a clear distinction between confidence and the aha experience. The overlap between these constructs arises from the language used to talk about insight. Descriptions typically used in the literature to describe an aha experience typically emphasize the "suddenness and obviousness" of the solution (e.g., Bowden & Jung-Beeman, 2003a; Danek et al., 2014a; Kizilirmak, Gomes da Silva, et al., 2016). The retrospective obviousness of the solution is arguably linked to a subjective increase in confidence. However, a high degree of confidence can arise from slower, analytic problem solving as well; consequently, the aha experience is distinguished from confidence in its suddenness. This dissociation could be tested using ratings of confidence and aha experience conditional on accuracy across trials: if surprise distinguishes confidence and the aha experience, then as solution accuracy becomes more reliable across trials, feelings of confidence will increase (e.g., Peirce & Jastrow, 1884; Yeung & Summerfield, 2012), and the aha experience will decrease. Our present methodology unfortunately does not enable us to make this distinction, since there was not sufficient control over the probable accuracy of response.

Performance on classic insight problems, compound remote associates, and anagrams was positively correlated, but not between classic noninsight problems and contemporary problems (Cinan, Özen, & Hampshire, 2013; Fleck, 2008; Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005; Wen et al., 2013; Wieth & Burns, 2000). This finding could reflect differences in the underlying processes of solving insight problems (i.e., restructuring). However, performance on classic insight and noninsight problems was also positively related (see also Gilhooly & Fioratou, 2009). This could reflect the similarity in the phrasing and presentation of the problems. Finally, performance on anagrams and compound remote associates was related, and again is likely to be due to similarities in their structure: both were short verbal problems requiring high crystalized intelligence and verbal fluidity. The absence of a relationship with accuracy on Raven's Advanced Progressive Matrices is consistent with this supposition.

Methodological implications

The present work raises several issues regarding the way insight problem solving is studied. A well-recognized yet pervasive issue in the literature regards the use of small numbers of tasks in an experiment (Bowden et al., 2005). For instance, 27 articles in the last decade have used a single insight problem to investigate individual differences in insight problem solving. The rationale for using small numbers of problems is clear; classic insight problems are highly diverse and have a low solution rate for any times less than 10 min (Bowden et al., 2005). However, the present research highlights the potential problems inherent in using a single classic problem as a test of insightful problem solving: There are large differences in accuracy and in reported insight affect among all problem tasks and types. One way to ameliorate these issues is to use contemporary problems, such as compound remote associates and anagrams, which allow for a larger number of problems to be tested in a given time period.

It is clear that insight problems, anagrams, and compound remote associates alike are able to elicit insight, and arguably both problem types require restructuring. However, it is important to note that compound remote associates and anagrams are distinctly different tasks from classic insight problems in their cognitive requirements. For example, verbal overshadowing hampers classic insight tasks (Schooler et al., 1993) but facilitates compound remote associates (Ball & Stevens, 2009). The present findings regarding the ability of compound remote associates and anagrams to elicit strong ratings of insight, particularly in the correct solution of the problem, reflects the fragmentation of methodology and findings arising from the different approaches to insight research, and reflect a need to consider once again what insight might mean; whether it is reflected by a feeling, task, or process.

Although normative data has been provided for many of these problems (e.g., Bowden & Jung-Beeman, 2003b), the data are predominately reaction times and solution rates. These are necessary statistics but given the rising interest in insight *in* problem solving and the lack of reliability of some problems in eliciting insight (e.g., Danek et al., 2016; Webb et al., 2016b), we offer this study both as an indicator for some problems in the literature and as a source to obtain problems that reliably elicit strong insight phenomenology.

Author note This work was supported by ARC Discovery Project Grant DP160102360 to Daniel R. Little.

Appendix

Table 4 Classic insight problems sorted in descending order according to aha elicited in correct solutions

Problem	Solved	Aha Correct			Aha Incorrect			RT to Solution	
	(%)	*mean*	*SD*	*skew*	*mean*	*SD*	*skew*	*mean*	*SD*
10 Tree	12	80.42	25.54	– 1.22	20.89	31.37	1.33	151.15	56.64
L Farm	16	76.94	26.73	– 1.16	43.92	36.56	0.04	150.51	39.17
Pig Farm	20	76.33	26.84	– 1.29	20.05	28.44	0.71	164.91	60.45
Marrying Man	28	74.93	21.76	– 1.31	27.83	27.76	0.91	122.50	50.66
Triangle	45	74.07	26.10	– 0.88	30.23	30.24	0.56	NA	52.20
Antique Coin	31	71.76	25.58	– 0.66	43.69	33.33	1.22	109.08	52.04
9 Dot	10	69.36	31.96	– 0.54	17.66	28.03	0.13	101.86	52.81
Hole	43	68.17	31.77	– 1.08	40.11	31.79	1.64	51.27	50.24
Socks	37	66.28	25.62	– 0.71	42.92	33.52	0.22	90.74	59.98
Train	28	65.79	28.69	– 0.65	38.03	28.39	0.24	132.52	56.74

 JrnlID 13428_ArtID 972_Proof# 1 - 25/09/2017

Table 4 (continued)

Problem	Solved	Aha Correct			Aha Incorrect			RT to Solution	
	(%)	*mean*	*SD*	*skew*	*mean*	*SD*	*skew*	*mean*	*SD*
Prisoner	58	65.51	32.91	– 0.57	26.67	26.42	0.92	116.90	41.85
Football	35	63.83	29.59	– 0.50	28.52	31.58	0.68		
Ladder	45	60.45	30.83	– 0.85	31.9	31.08	0.24	95.61	47.69
Horse Trader	56	59.23	32.37	– 0.45	46.89	31.12	0.58	76.99	57.77
Light	36	58.69	30.86	– 0.45	41.88	28.93	0.31	102.28	51.89
Pound Coins	23	58	39.52	– 0.52	41.75	31.68	0.15	86.92	56.76
Chain	07	57.29	37.06	– 0.31	24.67	24.10	0.93	175.64	57.50
Tumour	08	52.38	28.99	0.17	24.09	29.89	0.85	127.08	61.17
Ping Pong	23	48.91	32.97	0.09	45.96	28.19	0.35	62.57	60.50
Two String	16	47.47	30.97	– 0.24	40.31	29.80	– 0.02	109.81	55.46
Cherry	26	45.24	34.81	0.28	43.7	29.99	0.18	166.81	58.57
Candle	43	41.21	37.95	0.15	37.78	29.76	– 0.31	98.74	53.07
Card Hole	11	38.8	31.21	– 0.08	33.26	25.85	– 0.13	122.00	53.43
Pyramid	27	31.74	34.76	0.20	29.63	31.71	0.24	118.13	53.35
Lilies	57	30.44	31.74	– 0.06	76.36	29.20	0.54	85.21	59.34

Q29 **Table 5** Classic noninsight problems sorted in descending order according to aha elicited in correct solutions

Problem	Solved	Aha Correct			Aha Incorrect			Average RT	
	(%)	*mean*	*sd*	*skew*	*mean*	*sd*	*skew*	*mean*	*sd*
Fox Chicken	58	69.41	25.98	– 0.78	37.40	33.08	0.41	144.68	50.11
Water Jug	45	65.76	34.18	– 0.75	31.15	34.66	0.64	163.09	51.21
RA3	86	62.87	30.15	– 0.60	43.18	34.63	0.20	54.59	39.80
RA4	88	62.47	33.03	– 0.61	76.11	23.50	– 0.10	32.03	21.92
RA5	85	62.24	31.10	– 0.55	58.09	32.20	– 0.34	48.80	31.70
RA6	70	62.23	26.91	– 0.40	48.35	28.19	0.39	48.60	34.61
RA2	81	62.01	30.81	– 0.60	34.82	31.87	– 0.15	42.37	33.88
Hanoi	55	61.07	31.07	– 0.47	63.17	30.44	– 0.80	139.50	48.07
Calendar	80	60.57	35.65	– 0.56	50.60	32.58	– 0.04	105.95	47.68
RA1	94	60.33	30.10	– 0.62	32.17	28.82	– 0.07	40.85	29.25
RA7	66	59.39	31.61	– 0.59	51.64	24.77	– 0.33	54.75	32.16
RA12	43	58.51	29.63	– 0.43	46.25	31.59	– 0.09	79.44	56.68
Age	47	58.44	36.66	– 0.43	30.39	35.05	0.68	157.49	51.13
Flower	84	56.37	33.60	– 0.30	38.31	41.84	0.36	188.12	28.54
Police	33	54.36	34.68	– 0.22	45.81	29.62	– 0.10	142.01	53.51
RA9	49	53.35	29.66	– 0.38	49.57	29.95	0.01	70.13	40.61
RA8	58	52.38	31.48	– 0.24	46.78	30.51	– 0.24	80.38	55.85
Dinner	86	51.12	34.72	– 0.23	49.57	33.45	– 0.23	113.30	37.81
RA10	57	49.84	31.53	– 0.10	39.63	32.57	0.30	86.96	50.00
Cards	50	48.79	29.49	0.06	37.18	36.35	0.47	88.19	34.77
Weigh Coins	52	47.52	33.45	0.03	42.93	30.66	0.09	144.82	63.09
RA11	53	46.56	30.66	– 0.25	42.63	31.12	0.02	109.33	55.08
Hobbits	09	46.44	31.60	0.33	45.97	34.79	– 0.15	145.62	61.49
Puzzling Puzzle	72	43.26	31.12	0.14	37.00	28.95	0.26	108.87	51.91
Bachelor	44	40.70	31.62	0.26	23.22	32.76	1.09	198.72	25.98
ToL	38	39.77	27.66	0.59	33.16	32.08	0.23	185.92	36.22

Table 6 Compound remote associates sorted in descending order according to aha elicited in correct solutions

Problem	Solved	Aha Correct			Aha Incorrect			Average RT	
	(%)	*mean*	*sd*	*skew*	*mean*	*sd*	*skew*	*mean*	*sd*
sweet	20	87.43	23.29	–2.55	21.80	27.85	1.15	23.33	8.92
day	53	79.78	24.15	–1.38	29.15	34.30	0.83	17.36	9.97
care	44	78.51	26.64	–1.57	28.71	33.39	0.86	17.98	9.87
nut	36	77.86	31.03	–1.42	18.92	25.44	1.33	21.41	8.96
blue	52	77.83	26.66	–1.53	28.77	31.32	0.66	16.87	9.29
power	7	77.75	18.51	0.11	17.45	27.01	1.52	26.05	6.44
tape	64	77.17	25.94	–1.56	19.64	27.91	1.43	16.09	8.96
fire	75	77.03	27.33	–1.52	25.52	30.47	0.78	13.15	8.61
super	26	76.93	25.66	–1.13	27.12	31.18	0.77	21.69	8.75
gold	55	76.34	24.51	–1.17	24.24	27.49	0.94	17.84	9.29
fast	52	74.45	28.80	–1.35	18.35	27.21	1.35	18.66	9.90
stone	23	73.58	25.41	–0.75	18.29	27.44	1.50	23.80	8.23
match	17	73.50	29.77	–1.49	16.86	23.67	1.39	23.68	8.51
eye	9	73.30	36.02	–0.86	21.43	28.51	1.17	24.56	7.65
bag	76	72.56	29.22	–1.17	20.67	28.19	1.14	13.38	8.94
common	45	72.24	29.27	–1.24	23.64	25.60	0.85	18.82	9.25
birth	14	71.60	37.70	–0.93	22.79	28.55	1.02	23.90	7.81
book	56	71.51	28.74	–0.97	22.39	26.13	1.11	16.86	9.36
dead	19	71.35	27.48	–0.79	27.00	32.45	0.84	22.76	8.51
soap	28	70.00	28.75	–0.97	22.35	29.19	1.07	22.59	8.84
figure	19	69.50	29.84	–0.73	26.30	29.17	0.75	21.91	9.18
pin	51	69.29	28.97	–1.02	29.57	31.30	0.73	18.06	9.68
grand	14	67.93	31.64	–0.97	21.99	28.26	0.96	23.69	7.91
blind	20	67.90	29.88	–0.66	17.66	27.26	1.52	23.59	7.89
brain	10	66.82	29.58	–0.73	19.56	28.50	1.24	25.56	7.33
gun	5	66.50	28.17	–0.38	21.71	30.84	1.33	26.29	5.87
lip	6	66.29	32.67	–0.93	24.82	27.16	0.65	24.21	7.88
sun	27	66.00	29.05	–0.78	25.77	21.76	1.38	21.97	9.07
boy	43	65.00	28.40	–0.73	17.88	26.18	1.30	20.55	8.98
battle	16	63.35	35.13	–0.62	18.92	27.86	0.75	24.68	7.45
blood	27	62.11	33.12	–0.49	29.77	30.28	0.60	21.19	9.43
boat	28	60.38	28.99	–0.47	31.85	33.87	1.06	20.00	8.65
school	29	58.33	34.19	–0.29	23.24	28.21	0.22	22.91	8.06
star	62	52.33	33.45	–0.17	40.84	35.23	1.26	19.00	8.56

Table 7. Anagrams sorted in descending order according to aha elicited in correct solutions

Problem	Solved	Aha Correct			Aha Incorrect			Average RT	
	(%)	*mean*	*sd*	*skew*	*mean*	*sd*	*skew*	*mean*	*sd*
final	68	82.12	20.44	–1.71	20.38	31.38	1.40	16.94	9.82
joker	73	80.32	22.39	–1.64	29.89	38.49	0.69	15.90	9.86
pilot	70	79.24	25.06	–1.52	23.10	33.58	1.22	16.71	9.69
sound	85	78.60	24.57	–1.45	15.67	25.98	1.03	13.14	9.29
basic	81	78.16	24.88	–1.38	18.74	27.43	1.25	14.26	8.64
night	55	77.66	29.22	–1.40	62.64	33.51	–0.53	12.77	8.52

Table 7. (continued)

Problem	Solved	Aha Correct			Aha Incorrect			Average RT	
	(%)	*mean*	*sd*	*skew*	*mean*	*sd*	*skew*	*mean*	*sd*
piano	93	77.55	26.04	– 1.11	23.29	39.94	0.82	10.85	7.22
human	81	77.49	22.95	– 1.23	18.11	30.05	1.39	14.23	8.81
force	66	76.18	24.78	– 1.13	11.88	24.02	2.33	16.27	10.43
daisy	79	75.94	27.51	– 1.39	43.86	41.20	0.13	13.48	8.33
glove	69	75.74	26.33	– 1.27	23.52	33.45	1.04	16.26	9.78
joint	83	75.70	25.48	– 1.08	38.00	38.85	0.43	14.71	8.56
pouch	78	75.41	25.85	– 1.21	18.77	29.56	1.43	16.89	8.96
chair	89	75.16	26.12	– 1.04	41.55	44.36	0.39	10.08	7.19
party	50	74.96	25.68	– 0.94	21.14	34.34	1.40	20.72	9.62
frame	81	74.91	26.27	– 1.03	18.63	32.18	1.55	11.91	9.32
flour	93	74.68	25.62	– 1.02	63.43	30.11	– 0.19	9.06	5.78
blaze	42	74.65	28.24	– 1.30	16.00	26.63	1.83	20.12	10.23
beach	91	74.48	27.56	– 1.21	30.44	42.66	0.64	11.32	7.41
shore	87	74.40	27.31	– 1.26	25.08	34.73	0.98	14.74	8.07
brown	94	74.36	27.27	– 1.02	43.67	35.91	– 0.27	10.07	6.89
house	87	74.27	28.56	– 1.20	29.00	36.89	1.02	11.92	7.80
rough	93	73.67	27.40	– 1.14	56.00	45.38	– 0.18	9.16	6.13
place	72	73.64	26.05	– 1.17	14.64	28.00	1.81	14.41	10.39
stand	81	73.46	25.02	– 0.88	19.26	30.18	1.05	14.49	8.83
train	82	73.16	29.44	– 1.10	31.44	41.03	0.71	11.77	8.58
cloud	89	72.39	27.73	– 1.06	48.36	39.82	– 0.07	11.03	8.02
child	93	72.20	29.58	– 1.01	41.71	44.33	0.35	8.88	6.56
scout	89	71.68	27.50	– 1.06	71.27	34.39	– 0.81	10.50	6.51
chime	34	71.03	27.32	– 1.03	13.98	25.73	1.82	24.90	7.62
cruel	95	70.76	29.82	– 0.99	41.20	46.59	0.21	8.91	6.05
grant	76	70.36	29.01	– 0.89	19.04	31.10	1.30	16.50	8.99
ranch	85	68.93	30.27	– 0.85	20.00	25.28	0.92	14.22	9.04
hazel	65	66.30	31.91	– 0.85	33.69	38.51	0.52	15.17	9.65

References

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *54*, 394–403. doi:https://doi.org/10.1177/0013164494054002013

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

Aziz-Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009). "Aha!": The neural correlates of verbal insight solutions. *Human Brain Mapping*, *30*, 908–916. doi:https://doi.org/10.1002/hbm.20554

Ball, L., & Stevens, A. (2009). Evidence for a verbally-based analytic component to insight problem solving. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1060–1065). Austin: Cognitive Science Society. Retrieved from http://www.csjarchive.cogsci.rpi.edu/proceedings/2009/papers/251/paper251.pdf

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2014). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods*, 47: 918–929. doi:https://doi.org/10.3758/s13428-014-0530-7

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

Bowden, E. M., & Beeman, M. J. (1998). Getting the right idea: Activation in the right hemisphere may help solve insight problems. *Psychological Science*, *9*, 435–440.

Bowden, E. M., & Jung-Beeman, M. (2003a). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin & Review*, *10*, 730–737. doi:https://doi.org/10.3758/BF03196539

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

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

Chermahini, S. A., Hickendorff, M., & Hommel, B. (2012). Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach. *Thinking Skills and Creativity*, *7*, 177–186. doi:https://doi.org/10.1016/j.tsc.2012.02.003

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 14–27. doi:https://doi.org/10.1037/0278-7393.30.1.14

Cinan, S., Özen, G., & Hampshire, A. (2013). Confirmatory factor analysis on separability of planning and insight constructs. *Journal of Cognitive Psychology*, *25*, 7–23. doi:https://doi.org/10.1080/20445911.2012.729035

Cranford, E., & Moss, J. (2012). Is insight always the same? A protocol analysis of insight in compound remote associate problems. *Journal of Problem Solving*, *4*, 128–153. doi:https://doi.org/10.7771/1932-6246.1129

Cunningham, J. B., MacGregor, J. N., Gibb, J., & Haar, J. (2009). Categories of insight and their correlates: An exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *Journal of Creative Behavior*, *43*, 1966–1967.

Cushen, P. J., & Wiley, J. (2011). Aha! Voila! Eureka! Bilingualism and insightful problem solving. *Learning and Individual Differences*, *21*, 458–462. doi:https://doi.org/10.1016/j.lindif.2011.02.007

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

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*, 659–669. doi:https://doi.org/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*, 1–11. doi:https://doi.org/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*, 174–185. doi:https://doi.org/10.1016/j.cognition.2013.11.003

Danek, A. H., Wiley, J., & Öllinger, M. (2016). Solving classical insight problems without aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *Journal of Problem Solving*, *9*, 47–57.

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

DeCaro, M. S., Van Stockum, C. A., Jr, & Wieth, M. B. (2016). When higher working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 39–49. doi:https://doi.org/10.1037/xlm0000152

Derbentseva, N. (2007). *The intensity of the insight experience in problem solving: Structural and dynamic properties*. PhD dissertation, University of Waterloo, Waterloo, ON, Canada.

Dominowski, R. L., & Buyer, L. S. (2000). Retention of problem solutions: The re-solution effect. *American Journal of Psychology*, *113*, 249–274. doi:https://doi.org/10.2307/1423730

Duncker, K. (1945). On problem-solving (L. S. Lees, Trans.). *Psychological Monographs*, *58*(5, Whole No. 270), 1–113. doi:https://doi.org/10.1037/h0093599

Faber, A. (2012). *Investigation of insight with magic tricks: Introducing a novel paradigm*. PhD dissertation, Universität München, München, Germany.

Feynman, R. P. (1999). The pleasure of finding things out. In J. Robbins (Ed.). Cambridge: Perseus.

Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *European Journal of Cognitive Psychology*, *20*, 139–176. doi:https://doi.org/10.1080/09541440601016954

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*, 990–1006. doi:https://doi.org/10.3758/BF03196876

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: MIT Press.

Gilhooly, K., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking and Reasoning*, *15*, 355–376. doi:https://doi.org/10.1080/13546780903178615

Gilhooly, K., Fioratou, E., & Henretty, N. (2010). Verbalization and problem solving: Insight and spatial factors. *British Journal of Psychology*, *101*, 81–93. doi:https://doi.org/10.1348/000712609X422656

Gilhooly, K., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking and Reasoning*, *11*, 279–302. doi:https://doi.org/10.1080/13546780442000187

Guildford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1978). *Alternate uses: Manual of instructions and interpretation*. Orange: Sheridan Psychological Services.

Hill, G., & Kemp, S. M. (2016a). Heart rate differences in insight and non-insight problem solving. In *BPS CocSec*. doi:https://doi.org/10.1002/jocb.142.

Hill, G., & Kemp, S. M. (2016b). Uh-oh! What have we missed? A qualitative investigation into everyday insight experience. *Journal of Creative Behavior*. Advance online publication. doi:https://doi.org/10.1002/jocb.142

Jacobsen, R. (2016). Is eureka right? *Scientific American*, *315*, 18.

Jarman, M. S. (2014). Quantifying the qualitative: Measuring the insight experience. *Creativity Research Journal*, *26*, 276–288. doi:https://doi.org/10.1080/10400419.2014.929405

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1017–1027. doi:https://doi.org/10.1037/0278-7393.29.5.1017

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., … Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*, 500–510. doi:https://doi.org/10.1371/journal.pbio.0020097

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

Kizilirmak, J. M., Thuerich, H., Folta-Schoofs, K., Schott, B. H., & Richardson-Klavehn, A. (2016). Neural correlates of learning from induced insight: A case for reward-based episodic encoding. *Frontiers in Psychology*, *7*, 1–16. doi:https://doi.org/10.3389/fpsyg.2016.01693

Klein, G., & Jarosz, A. (2011). A naturalistic study of insight. *Journal of Cognitive Engineering and Decision Making*, *5*, 335–351. doi:https://doi.org/10.1177/1555343411427013

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*, 1534–1555. doi:https://doi.org/10.1037/0278-7393.25.6.1534

Köhler, W. (1921). *Intelligenzprüfungen am Menschenaffen*. Berlin: Springer. doi:https://doi.org/10.1007/978-3-642-47574-0

Kounios, J., & Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, *18*, 210–216. doi:https://doi.org/10.1111/j.1467-8721.2009.01638.x

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of memory (Book review). *Annual Review of Psychology*, *16*, 190–191. doi:https://doi.org/10.1097/00146965-200309000-00007

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., & Jung-Beeman, M. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia*, *46*, 281–291. doi: https://doi.org/10.1016/j.neuropsychologia.2007.07.013

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K. K., Parrish, T. B., & Jung-Beeman, M. (2006). The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychological Science*, *17*, 882–890. doi:https://doi.org/10.1111/j.1467-9280.2006.01798.x

Lezak, M. D. (Ed.). (2004). *Neuropsychological assessment*. New York: Oxford University Press.

Liljedahl, P. G. (2004). *The aha! experience: Mathematical contexts, pedagogical implications*. PhD dissertation, Simon Fraser University, Burnaby, BC, Canada.

Liljedahl, P. G. (2005). AHA!: The effect & affect of mathematical discovery on undergraduate students. *International Journal of Mathematical Education in Science and Technology*, *36*, 219–234.

Lin, W.-L., Hsu, K.-Y., Chen, H.-C., & Wang, J.-W. (2012). The relations of gender and personality traits on different creativities: A dual-process theory account. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 112–123. doi:https://doi.org/10.1037/a0026241

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*, 36–44. doi:https://doi.org/10.3389/fpsyg.2014.00239

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., … Morey, R. D. Wagenmakers, E.-J. (2015). JASP (Version 7) [Computer software].

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

MacGregor, J. N., & Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behavior Research Methods*, *40*, 263–268. doi:https://doi.org/10.3758/BRM.40.1.263

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*, 176–201. doi:https://doi.org/10.1037/0278-7393.27.1.176

Maier, N. R. F. (1931). Reasoning and learning. *Psychological Review*, *38*, 332–346. doi:https://doi.org/10.1037/h0069991

Mason, O. J., & Claridge, G. (2006). The Oxford–Liverpool Inventory of Feelings and Experiences (O-LIFE): Further description and extended norms. *Schizophrenia Research*, *82*, 203–211. doi:https://doi.org/10.1016/j.schres.2005.12.845

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232. doi:https://doi.org/10.1037/h0048850

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 288–294. doi:https://doi.org/10.1037/0278-7393.12.2.288

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

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

Murray, M. A., & Byrne, R. M. J. (2005). Attention and working memory in insight problem-solving. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (Vol. 27, pp. 1571–1575). Austin: Cognitive Science Society. Retrieved from www.psych.unito.it/csc/cogsci05/frame/poster/2/f285-murray.pdf

Nęcka, E., Żak, P., & Gruszka, A. (2016). Insightful imagery is related to working memory updating. *Frontiers in Psychology*, *7*, 137:1–11. doi:https://doi.org/10.3389/fpsyg.2016.00137

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

Ohlsson, S. (1984). Restructuring revisited. *Scandinavian Journal of Psychology*, *25*, 65–78.

Öllinger, M., Jones, G., Faber, A. H., & Knoblich, G. (2012). 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*, 931–939. doi:https://doi.org/10.1037/a0029194

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

Öllinger, M., & Knoblich, G. (2009). Psychological research on insight problem solving. In H. Atmanspacher & H. Primas (Eds.), *Recasting reality: Wolfgang Pauli's philosophical ideas and contemporary science* (pp. 275–300). Berlin: Springer.

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*, 791–799. doi:https://doi.org/10.1037/0278-7393.28.4.791

Ovington, L. A., Saliba, A. J., Moran, C. C., Goldring, J., & MacDonald, J. B. (2015). Do people really have insights in the shower? The when, where and who of the aha! moment. *Journal of Creative Behavior*. Advance online publication. doi:https://doi.org/10.1002/jocb.126

Paulewicz, B., Chuderski, A., & Nęcka, E. (2007). Insight problem solving, fluid intelligence, and executive control: A structural equation modeling approach. In *Proceedings of the 2nd European Cognitive Science Conference*. Retrieved from www.researchgate.net/publication/252109497_Insight_problem_solving_fluid_intelligence_and_executive_control_A_structural_equation_modeling_approach/file/e0b4951f28bf772fb8.pdf

Peirce, C. S., & Jastrow, J. (1884). On small differences of sensation. *Proceedings of the National Academy of Sciences*, *3*, 75–83.

Poincaré, H. (1913). *The foundations of science*. Lancaster: Science Press.

Qiu, J., Li, H., Jou, J., Wu, Z., & Zhang, Q. (2008). Spatiotemporal cortical activation underlies mental preparation for successful riddle solving: An event-related potential study. *Experimental Brain Research*, *186*, 629–634. doi:https://doi.org/10.1007/s00221-008-1270-7 Q33

Qualtrics. (2016). Qualtrics. Provo, Utah, USA. Retrieved from https://www.qualtrics.com/

Raven, J. C. (1985). *A Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis.

Raven, J. C. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*, 1–48. doi:https://doi.org/10.1006/cogp.1999.0735

Salvi, C., Bricolo, E., Bowden, E., Kounios, J., & Beeman, M. (2016). Insight solutions are correct more often than those achieved by analysis. *Thinking and Reasoning*, *22*, 443–460. doi:https://doi.org/10.1080/13546783.2016.1141798

Sandkühler, S., & Bhattacharya, J. (2011). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS One*, *3*, e1459. doi:https://doi.org/10.1371/journal.pone.0001459

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*, 166–183. doi:https://doi.org/10.1037/0096-3445.122.2.166

Schultz, G. (1890). Feier der Deutschen Chemischen Gesellschaft zu Ehren August Kekulé's. *Berichte der Deutschen Chemischen Gesellschaft*, *23*, 1265–1312.

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the "Aha!" experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology*, *107*, 281–298. doi:https://doi.org/10.1111/bjop.12142

Sternberg, R. J., & Davidson, J. E. (1995). *The nature of insight*. Cambridge: MIT Press.

Subramaniam, K., Kounios, J., Parrish, T. B., & Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *Journal of Cognitive Neuroscience*, *21*, 415–432. doi:https://doi.org/10.1162/jocn.2009.21057

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology*, *58A*, 1311–1323. doi:https://doi.org/10.1080/02724980443000593

Thevenot, C., & Oakhill, J. (2008). A generalization of the representational change theory from insight to non-insight problems: The case of arithmetic word problems. *Acta Psychologica*, *129*, 315–324. doi:https://doi.org/10.1016/j.actpsy.2008.08.008

Topolinski, S., & Reber, R. (2010). Gaining insight into the "aha" experience. *Current Directions in Psychological Science*, *19*, 402–405. doi:https://doi.org/10.1177/0963721410388803

Webb, M. E., Little, D. R., & Cropper, S. J. (2016a). *Chasing insight: Feelings of insight in the solution of insight and non-insight problems*. Paper presented at the 2016 annual meeting of the Australasian Society for Experimental Psychology. Melbourne, Australia.

Webb, M. E., Little, D. R., & Cropper, S. J. (2016b). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in Psychology*, *7*, 1424:1–13. doi:https://doi.org/10.3389/fpsyg.2016.01424

Webb, M. E., Little, D. R., Cropper, S. J., & Roze, K. (2017). The contributions of convergent thinking, divergent thinking, and schizotypy to solving insight and non-insight problems. *Thinking and Reasoning*, *23*, 235–258. doi:https://doi.org/10.1080/13546783.2017.1295105

Webb, M. E., Little, D. R., Cropper, S. J., & Webb, M. E. (2017). *Suddenly I see: Presentation of a solution enhances "aha" experience, particularly when the solution is unexpected*. Paper presented at the 2017 annual meeting of the Australasian Society for Experimental Psychology,.Shoal Bay, Australia.

Wegbreit, E., Suzuki, S., Grabowecky, M., Kounios, J., & Beeman, M. (2012). Visual attention modulates insight versus analytic solving of verbal problems. *Journal of Problem Solving*, *4*, 94–115. doi:https://doi.org/10.7771/1932-6246.1127

Wei, T., & Simko, V. (2016). Package "corrplot." Retrieved from https://github.com/taiyun/corrplot

Weisberg, R. W. (1995a). Case studies of creative thinking: Reproduction versus restructuring in the real world. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 53–72). Cambridge: MIT Press, Bradford Books.

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

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General*, *110*, 169–192. doi:https://doi.org/10.1037/0096-3445.110.2.169

Wen, M. C., Butler, L. T., & Koutstaal, W. (2013). Improving insight and non-insight problem solving with brief interventions. *British Journal of Psychology*, *104*, 97–118. doi:https://doi.org/10.1111/j.2044-8295.2012.02107.x

Wieth, M., & Burns, B. D. (2000). Motivation in insight versus incremental problem solving. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Meeting of the Cognitive Science Society* (pp. 550–564). Mahwah: Erlbaum.

Wieth, M., & Zacks, R. T. (2011). Time of day effects on problem solving: When the non-optimal is optimal. *Thinking and Reasoning*, *17*, 387–401. doi:https://doi.org/10.1080/13546783.2011.625663

Wiseman, R., Watt, C., Gilhooly, K., & Georgiou, G. (2011). Creativity and ease of ambiguous figural reversal. *British Journal of Psychology*, *102*, 615–622. doi:https://doi.org/10.1111/j.2044-8295.2011.02031.x

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *367*, 1310–1321. doi:https://doi.org/10.1098/rstb.2011.0416