The Eureka Heuristic: Relying on insight to appraise the quality of ideas

Ruben E. Laukkonen[1], Jonathan W. Schooler[2], & Jason M. Tangen[1]

r.laukkonen@uq.edu.au[1], jonathan.schooler@psych.ucsb.edu[2], jtangen@psy.uq.edu.au[1]

[1]Department of Psychology, University of Queensland, St Lucia, Queensland, Australia 4072

[2]Psychological & Brain Sciences, University of California, Santa Barbara, California, USA 93106

Corresponding author: Ruben Laukkonen, ruben.laukkonen@gmail.com

Abstract

Perhaps it is no accident that "Eureka" moments accompany some of humanity's most important discoveries in science, medicine, and art. This paper describes an account where insight experiences play an adaptive role, by aiding humans to choose the right solution to a problem. Experiments reveal that feelings of insight—without any conscious verification or deliberation—predict confidence and accurate solutions to problems. There is also evidence that humans self interpret their Aha! experiences. One possibility is that humans use insight phenomenology heuristically in order to appraise their own ideas. This functional view of insight speaks to a number of open questions in the literature: Why do insight experiences occur in certain contexts but not others? Why do insight experiences predict confidence and objective performance in some contexts but not others? Why are some insights more intense than others? What leads to false insights? We also propose the *insight fallacy* to describe situations where a person incorrectly concludes that a solution or idea must be true solely based on the fact that it was accompanied by an insight experience.

Keywords: Insight, Problem Solving, Heuristics and Biases, Cognition, Appraisal

Introduction

“Eureka!, I know the answer!” said Archimedes. “But how do you know?!” questioned Socrates. “Well, because I experienced a Eureka moment,” he replied.

In the fictitious exchange above, Archimedes has an idea that occurred to him suddenly, and he experienced a Eureka moment. When asked to justify his sudden confidence in a solution he had only just discovered, he recursively points back to the Eureka moment itself. While popular accounts of Eureka moments and insight experiences are rarely this self-aware, it is possible to detect hints of a similar deductive process. In 2012, while waiting to go to a concert, mathematician Yitang Zhang discovered the solution to the twin prime problem. He said that he “...immediately knew that it would work,” and then it took several months to verify his solution (Nisbett, 2015). The 19th century polymath Henri Poincaré was immediately certain about the accuracy of a solution that came to him when stepping onto an omnibus, and mathematician Jacques Hadamard said that, “on being very abruptly awakened by an external noise, a solution long searched for appeared to me at once without the slightest instant of reflection on my part.”

In each of the above examples, the individual arrives at a sudden and unexpected solution to a problem, and is immediately certain that it is correct without deliberately checking. The same experience is regularly observed in the laboratory. Participants are more confident about solutions that are accompanied by insight experiences (Danek et al., 2014; Danek & Wiley, 2017; Webb et al., 2016), and insight experiences can strike suddenly and unexpectedly, even while engaged in another task (Ovington et al., 2015; Snyder, Mitchell, & Ellwood, 2004). If a result appears in mind unexpectedly, or while engaged in another task, then the solution was a product of processing that occurred below awareness. The fact that people are largely unaware of the problem solving processes that underlie their final solution

is also corroborated by numerous experiments. Maier (1931) famously found that participants suddenly discovered a solution to his 'two-string' problem shortly after he provided a hint that the participants claimed not to notice (an effect replicated by, Landrum, 1990). Bowden (1997) found that subliminally priming the solution to an anagram lead to more reported insight experiences, without participants knowing that they were being primed. Hattori, Sloman, and Orita (2013) also found that subliminal primes improved insight problem solving across three experiments, in some cases leading to a fivefold improvement.

Since the problem solving process—and therefore the reasoning that underlies the solution—is not directly available for introspection, what is the source of such an immediate sense of certainty? Perhaps it is the sudden insight phenomenology itself that compels humans to trust that the present solution—the only solution that evoked such intense feelings—is the one they have been searching for. The processing that precedes an insight solution may draw on a vast and complex network of information, experiences, and beliefs that are difficult and inefficient to consciously appraise. Time pressures also discourage deliberate evaluation. When a solution ultimately emerges, one may not have the luxury of weighing up its components before acting. Therefore it is often less important to know 'why you know,' and is abundantly more efficient to use feeling to signal that a viable solution has been found (Goldstein & Gigerenzer, 2002). It is plausible that the feeling of insight serves as a signal that a promising solution has been discovered through unconscious processes, and that humans rely on insight as a fast and frugal substitute for an effortful and time-consuming review of the evidence. In this paper, we describe how this view of insight experiences is consistent with the evidence, and affords many novel directions for future research.

As a substitute for analytic processing, the insight experience certainly appears to be functional. Solutions accompanied by insight experiences are more likely to be correct than

non-insight solutions (Danek et al., 2014; Danek & Wiley, 2017; Hedne et al., 2016; Salvi et al., 2016; Webb, Little, & Cropper; 2016). In one such example, Laukkonen, Ingledew, Schooler, and Tangen (2018) measured insight phenomenology in real-time using a highly sensitive measure of grip strength. Participants were instructed to squeeze the dynamometer as soon as they felt an insight experience occurring. They found that insight experiences—prior to any reflection on the solution by the participants—were substantially more likely to be correct for problems that involve implicit processing. In other words, the phenomenology of insight alone was sufficient to predict the accuracy of the response. The results were extended to a multisensory identification task, a paradigm highly representative of the kinds of identifications made in everyday life (i.e., naming familiar songs, faces, and aromas). Beyond the mere presence of insight, in both experiments, the intensity of the insight experience was a further positive predictor of accuracy. The same basic result has now been replicated in at least six studies with large effect sizes. How can insight phenomenology, and its intensity, predict accurate solutions to problems?

It is common knowledge that an expert can have intuitive expertise about their domain, for example the next best move on a chess board (Ericsson & Charness, 1994; Kahneman, 2015). Since humans often do not have direct access to their own problem solving processes, it's plausible that they also have intuitions about the accuracy of their ideas. Expert intuitions are often fast and feel automatic, and so too an intuition about the quality of an idea might occur immediately as the idea 'pops into mind.' A possible mechanism for this effect is that the insight phenomenology signals a consistency or coherence with what one knows and believes. Thus in the same way that the chess expert draws on her expertise to make a move—often without any conscious effort—the problem solver can evaluate a solution automatically and intuitively based on her own expertise

regarding the problem domain. This mechanism may explain why there is a positive relationship between insight experiences and accuracy. As long as a person's experiences are reliable, the greater the consistency between the solution and experience, the more likely the solution is accurate. An accuracy advantage for the insight experience would also be constrained to situations where some unconscious processing is involved (as found in Laukkonen et al., 2018). If there is no unconscious processing, then there is no intuition to be had. Likewise, if a novice has no experiences in a domain, then her intuitions will be of little use. A schematic of the *Eureka Heuristic* model is provided in Figure 1.

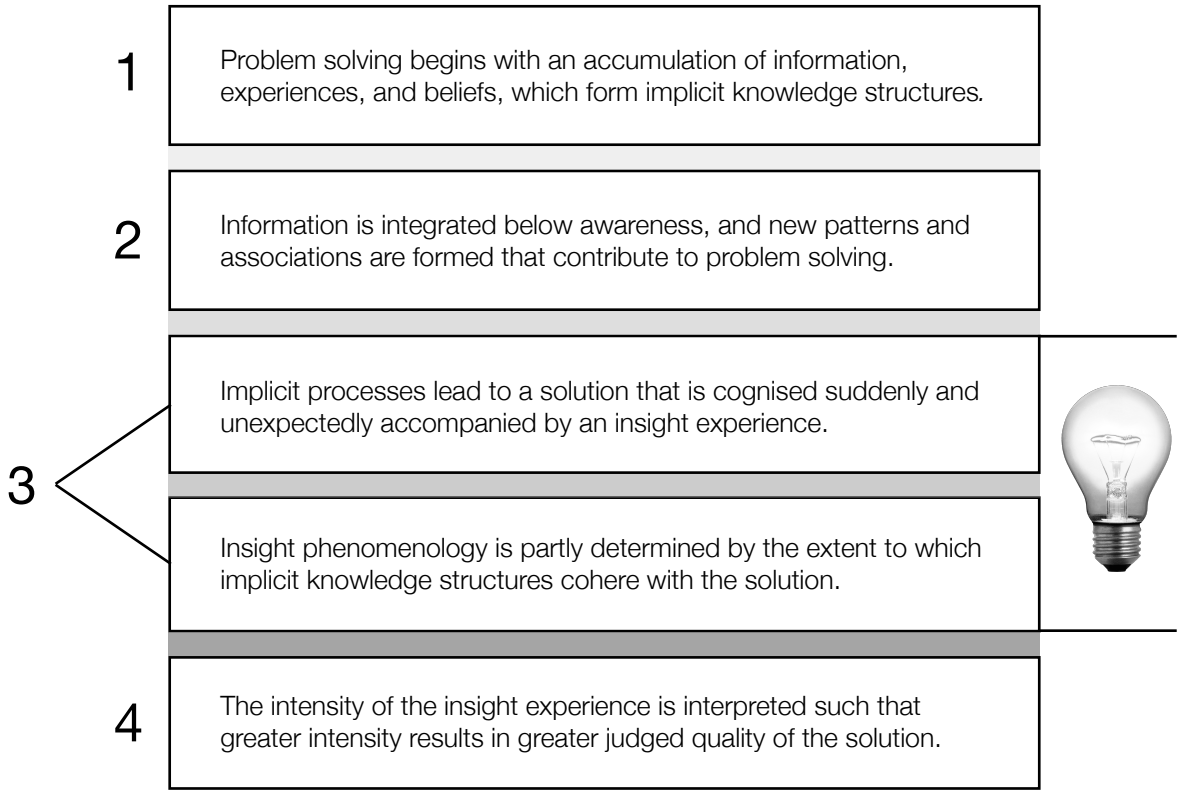

Figure 1. The four steps in the Eureka Heuristic view of the insight experience.

The heuristic view also helps to explain why insight experiences can range from being barely noticeable to intense, and why they can occur in a such a wide variety of circumstances. Mathematicians such as Henry Poincaré and Yitang Zhang, as well as physicists Albert Einstein and Richard Feynman, report intense Eureka experiences that resulted in, as Poincaré put it, "immediate certainty." Their considerable expertise, which was

developed over decades of study and practice resonated loudly with the sudden solution. On the other hand, in problems devised for lab-based experiments where one has minimal relevant experience, smaller insights tend to occur. Although less common, it also happens that one can have an insight experience where the contents of the solution are untrue (Danek & Wiley, 2017). If a person's knowledge or understanding is impoverished, then false insights are likely to arise. On the other hand, if one's memory, conceptual structures, and assimilated knowledge are based on decades of quality experience and clear and replicable evidence (and a healthy state of mind), then the Eureka moment can signal a breakthrough discovery.

We begin by describing the phenomenology of insight in the context of problem solving and discuss Topolinski & Reber's (2011) fluency account. We then outline the connection to self-interpretation, and discuss why it is appropriate to view the insight experience within the context of heuristics and biases. We also consider potential trade-offs involved with the Eureka heuristic, and describe the *insight fallacy* as any situation where one concludes that an idea is correct solely on the grounds that it was accompanied by insight phenomenology. In the final section, we discuss how the Eureka heuristic model may contribute to a number of long-standing debates in the insight problem solving literature, and provide a framework for understanding recent empirical findings. Given the breadth of research discussed, it is inevitable that some of the richness of each area is lost. We see that part of the value of this contribution is in describing overlapping literatures that are otherwise isolated, so as to provide novel perspective on an elusive phenomenon.

The Feeling of Insight

Recent definitions of insight emphasize the phenomenology that accompanies some sudden solutions such as pleasure, relief, drive, surprise, and in particular a sense of immediate obviousness of a solution (Danek & Wiley, 2017; Webb et al., 2016). Although the insight experience is more common under certain circumstances—particularly in creative problems that involve unconscious processing or representational change—it is not strictly problem-specific, nor is it ever guaranteed (Webb et al., 2016). Thus, over time, it has become common practice to measure when insight occurs on a case-by-case basis, and also to use a single set of problems to investigate both insight and non-insight solutions classified according to self-report (Bowden & Jung-Beeman, 2007). Given this state of affairs, it's surprising that theoretical contributions regarding the phenomenology of insight are rare. This lack of theory has become problematic recently because the phenomenology is increasingly used to define insight, and self-reported insights are used as dependent measures (Bowden & Jung-Beeman, 2007; Kounios & Beeman, 2014).

To the best of our knowledge, the fluency account of insight is currently the only published explanation of the insight phenomenology, and appears to be gaining popularity (Topolinski & Reber, 2010). According to Topolinski and Reber (2010), when an unexpected solution appears in mind during an Aha! moment, a problem that was once difficult or confusing is suddenly resolved and processed fluently, leading to positive affect and judged truth. Studies find that manipulating fluency creates a sense of cognitive ease, pleasure, and confidence (Topolinski & Strack, 2009; Winkielman & Cacioppo, 2001). For example, statements that are presented with high figure-ground contrast (e.g., black letters with a white background) are more likely to be judged as true compared to low contrast statements (Reber & Schwarz, 1999). Repeated exposure makes a stimulus more pleasurable (Reber,

Winkielman, & Schwarz, 1998), and solutions presented more 'suddenly' following an anagram (50ms versus 150ms) are more likely to be judged as correct (Schwarz, Newman, & Leach, 2016). The authors argue that since suddenness, pleasure, and judged truth are dimensions of insight, then fluency is likely to be the driving force behind the experience of insight (Topolinski & Reber, 2010).

It's important to consider what the fluency description of insight affords us, which wasn't previously known. It is known, for example, that the insight experience is associated with greater confidence or presumed accuracy of a solution that appears in mind, but it is *not* known why or how. According to Topolinski and Reber (2010), it is due to the experience of fluency that is inherent to the insight phenomenology, which has previously been shown to predict judged truth. However, this raises the question, *why* does fluency lead to increased confidence or judged truth? With regard to the open questions extant in the insight literature, we risk simply passing the buck from the feeling of insight to the feeling of fluency. Moreover, the fluency account—even if true—does not speak directly to any of the questions outlined at the beginning of this paper (e.g., why do insight experiences predict objective performance in some contexts but not others? Why are some insights more intense than others? What leads to false insights?). Fluency does make promising predictions about situations that might elicit *illusions* of insight, for example by artificially increasing the fluency at the moment of solution. But it seems improbable that the myriad of false insights humans have—especially those associated with complex belief systems—are driven by incidental states of fluency.

There are also quintessential elements of the insight experience not accounted for by fluency. According to participants' own reports, relief is a key feature of the insight phenomenology not connected to fluency (Danek & Wiley, 2017). Another dimension is drive, or inspiration, which accompanies some insight experiences (Danek & Wiley, 2017).

Archimedes was said to be so stimulated and inspired that he ran naked through the streets shouting “Eureka!”, pointing to the archetypal role of inspiration and the ‘rush of insight’ that accompanies the insight experience (Gick & Lockhart, 1995). Fluency also cannot easily explain why an insight can be barely noticeable in some cases, and phenomenally large in others. Andrew Wiles describes his discovery of the solution to Fermat’s last theorem in 1994 as follows:

> “At the beginning of September I was sitting here at this desk when suddenly, totally unexpectedly, I had this incredible revelation. It was the most important moment of my working life. Nothing I ever do again will... I’m sorry.”

Andrew Wiles fights back tears throughout the video, and in the end turns away from the camera because recounting the experience evokes a powerful emotional response. There is an apparent incongruity between the sheer emotional weight of some insight moments and the effects we observe (or would expect) from changes in fluency.

Our overall impression is that fluency is a parsimonious description of certain features of insight phenomenology. However, fluency does not fully account for the dimensions or the intensity of the insight experience. It is also quite clear that many key questions with regard to the behavioral consequences of the insight experience remain unanswered. It may also be that pigeonholing insight experiences as another instance of fluency may inadvertently lead to omitting the nuance of insight, and therefore to overlooking the unique role that it plays in decision-making.

Self Interpretation

In this section, we begin to unpack the studies that provide the scaffolding for the Eureka heuristic. One assumption of our account is that self interpretation is a basic feature of human cognition, one that is very likely exploited in the case of insight experiences. Participants often cannot reliably introspect regarding their mental processes or the true causes that underlie their behavior and attitudes, and regularly confabulate instead (Brasil-Neto et al., 1992; Johansson et al., 2004; Schooler, 2002; Wegner & Wheatley, 1999; Wegner, 2002). This is not to say that introspection is always incorrect, but that so-called introspections even when they are accurate, are in fact post-hoc interpretations based on implicit causal theories and an evaluation of contextual and sensory information. There are dozens of experiments that find a mismatch between self-reported causes of behavior (and underlying cognitive processes) and the actual causes triggered by various manipulations (see Nisbett & Schachter, 1966; Latane & Darley, 1970 for famous examples, Nisbett & Wilson, 1997 and Caruthers, 2009 for reviews, and for more recent work see Dougal & Schooler, 2007, Johansson et al., 2004; Johansson et al., 2005; Johansson et al., 2006; Whittlesea & Williams, 1998).

The classic research conducted by Maier (1931) provides an illustrative and relevant example. Maier (1931) set up an insight problem using two ropes hanging from the ceiling. The ropes each had different objects attached at the bottom, such as pliers, or clamps. The task was to attach the two ropes, but it was physically impossible to reach one rope while holding the other. When participants were close to giving up, Maier would haphazardly set one of the ropes in motion. Shortly thereafter participants tended to ‘suddenly’ discover the solution: they tied an object to the rope, set it in motion, quickly grabbed the other rope and then caught the swinging object and tied the ropes together. When probed about the source

of the solution, individuals confabulated that it simply ‘dawned on them,’ or that it was ‘the next obvious thing to try’ (for a replication, see Landrum, 1990).

Misattributions of phenomenology highlight how people interpret their feelings, despite the fact that the source of those feelings is hidden. In one such study, Dougal and Schooler (2007) presented participants with 60 words to memorize, and then provided a set of anagrams to solve followed by a recognition judgment regarding the solution of the anagram. They found that the anagrams that were solved were more likely to be recognised compared to the anagrams that were not, suggesting that something about solving the anagram was leading to an ‘illusion of prior experience.’ In five more experiments, Dougal and Schooler (2007) replicate their basic finding and argue that participants seem to fall prey to ‘discovery misattribution’: The Aha! experience of solving an anagram leads to a false inference of remembering, where participants incorrectly interpret their insight as a signal that a word is familiar. Not only is this study a further example of self-interpretation, the authors specifically show an effect of self-interpretation with regard to the insight experience.

In a series of similar experiments conducted by Whittlesea and colleagues (1990; 1998; 2000), they suggest that the feeling of surprise—a dimension of the insight experience—can also confound memory judgments, where pseudohomophones of real words (e.g., frog spelled *phrawg*) are more likely to be reported as old (recognised as previously seen) than words spelled correctly and non-words. The surprise experienced as a consequence of an unfamiliar letter-string, which sounds like a real word may have lead to a misattribution of phenomenology so that participants felt that the word was previously seen (Whittlesea & Williams, 2000).

Based on these data, we may conclude that the insight experience and some of its dimensions are self-interpreted in ways that lead to incorrect judgments in certain contexts (Dougal & Schooler, 2007; Whittlesea & Williams, 2000). We see no reason to believe that insight phenomenology is self-interpreted in these contexts, but not in the context where they most commonly occur: problem solving. It would be particularly surprising given that problem solving is precisely where the self-interpretation of insight experiences would be functional, given that they predict objective performance (Salvi et al., 2016).

Studies such as Maier (1930) show that individuals can fail to explain how they arrived at an insight solution. If people are self interpreting their insight moments then an additional prediction is that they may also fail to explain *why* an insight solution is correct. In many cases where self interpretation and confabulation occurs, people believe that they are genuinely introspecting (Caruthers, 2009). Hence, when a sudden insight occurs, even if it is possible to provide a narrative about why the solution is correct, it too may be a matter of post-hoc theorising and self interpretation. One prediction is that it ought to be possible to have an insight experience and provide an accurate solution, but also provide an inaccurate description about why that solution is correct. The insight phenomenology may in some cases be more trustworthy than one's own deliberate rationalising (Gigerenzer & Gaissmaier, 2011).

The Eureka Heuristic

In this section, we aim to describe how the insight experience is best understood as a heuristic, and why this view is a plausible interpretation of existing data. The heuristics and biases approach has had an enormous impact on decision-making research over the past 40 years (Simon, 1956; 1982; Tversky & Kahneman, 1973; 1975; Gigerenzer & Gaissmaier,

2011). The majority of this progress in understanding how humans make decisions comes from a deceptively simple idea that there are shortcuts to navigating a complex world. Stereotyping is a familiar example where humans categorise people and objects according to the sum of their experiences with them. Due to the inherent fact that humans have limited exposure to the members of any category—and limited cognitive capacities—they are forced to generalize from the small subset that they have been exposed to. They rely on a small subset (stereotype) to make predictions about new instances, which is an adaptive mental shortcut because it *usually* works.

For the purposes of considering the insight experience within this framework, the affect heuristic affords a particularly useful analogy. Slovic et al., (2007) pointed out that, "Although analysis is certainly important in some decision-making circumstances, reliance on affect and emotion is a quicker, easier, and more efficient way to navigate in a complex, uncertain, and sometimes dangerous world." They also cite an excerpt from Damasio (1994) who studied patients who suffered brain damage in ventromedial frontal cortices, which resulted in a specific impairment in the ability to "feel." Counterintuitively, the patients experienced a marked failure to make rational decisions and to reason appropriately, despite appearing to have all their other reasoning faculties intact. Slovic et al., (2007) suggest that affective cues are based on impressions developed through experience, where some object or event has been associated with positive or negative affect in the past. In a new uncertain situation, a person can draw on her impressions or experiences of similar situations by consulting her affective responses as a heuristic. Studies show that affect has widespread influence on judgments and decisions even in abstract domains that seem on the surface purely analytical, including risk judgments (Fischhoff et al., 1978), gambling and probability

judgments (Lowenstein et al., 2001), and a range of preference evaluations (Anderson, 1981; Mellers et al., 1992; Winkielman, Zajonk, & Schwarz., 1997).

The analogy between the affect heuristic and the Eureka heuristic has many levels. One obvious similarity is that they both involve an interpretation of phenomenology to guide decisions. A crucial difference lies in the underlying mechanisms. Slovic et al., (2007) proposed that the affective response to a stimulus or situation draws on an "affect pool" that contains previous, related experiences and representations tagged as either emotionally positive or negative. When in a familiar situation, a person can draw on the affect pool by reacting to the affect they currently experience to help guide decision-making. The Eureka heuristic may function in a similar way—where the affect heuristic draws on an "affect pool," the Eureka heuristic draws on a "knowledge pool." When a solution is found, and existing knowledge about the problem and relevant associations cohere with the new solution, then an insight moment occurs to signal that one's complex, implicit knowledge structures are consistent with the current solution. Another way to conceptualize this difference is that intuitions and affective cues occur to inform us about *events in the world*, whereas the insight experience occurs to inform us about *events in our minds* (see Figure 2). The same mechanisms that drive our intuitions about the world may therefore also drive our insight experiences.

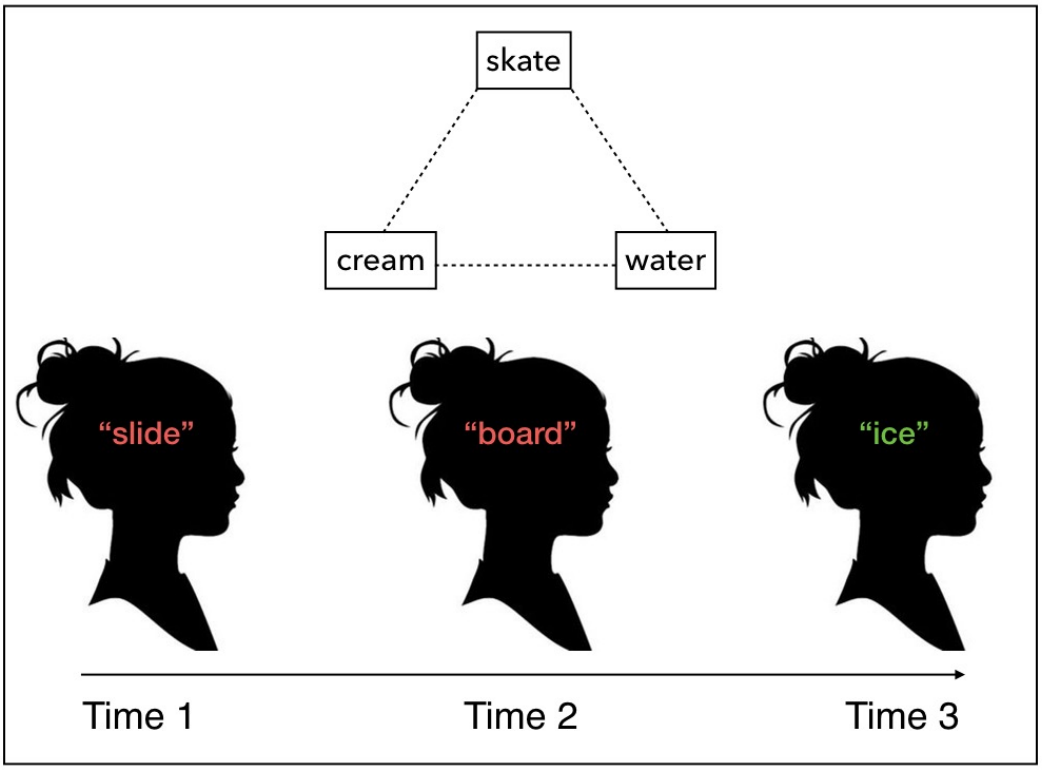

Figure 2. The participant is presented with some problem (e.g., a compound remote associate). Two solutions come to mind: “slide” at Time 1, and “board” at Time 2, but neither solution is compatible with all three words. At Time 3, the word “ice” appears in mind along with an insight experience, so she infers that no further processing is required and reports the solution.

If insight moments are used as an ecologically rational heuristic akin to recognition, or affect (Gigerenzer & Gaissmaier, 2011; Schooler & Hertwig, 2005; Slovic et al., 2007), then this account ought to predict problem solving accuracy and subjective confidence in the solution. The data are clear on both fronts. Across a range of problems, solutions that are accompanied by the insight experience are more likely to be accurate than solutions that are not, and insight moments predict confidence—despite no deliberate verification or introspection by the problem solver (Danek et al., 2014; Danek & Wiley, 2017; Hedne et al., 2016; Laukkonen et al., 2018; Metcalfe & Wiebe, 1987; Salvi et al., 2016; Webb, Little, & Cropper; 2016). Importantly, the relationship between insight and accuracy varies depending on the problem type.

A multiplication problem like 46 x 83 is rarely solved unconsciously through associative processing or restructuring, so one is unlikely to experience an insight moment. This is likely why, for classic analytical problems, there are fewer insight moments, and the correlation between insight and accuracy is negligible or non-existent (Danek et al., 2017; Laukkonen et al., 2018; Webb et al., 2016). The more the problem is amenable to a linear, conscious strategy, the more easily one can articulate the basis for the solution, and there is little use for informative phenomenology. Consistent with this view, in a comparison of the think aloud protocols leading up to successful solutions of problems associated with insight versus non-insight solutions, Schooler and Melcher (1995) found that numerous elements (e.g., arguments, re-reads) predicted analytic problem solutions, but very few predicted insight solutions. Metcalfe and Wiebe (1987) also showed that feelings of progress on a problem predict solutions for analytic but not insight problems. If the solution to a problem is simply the final step on a staircase then there's no need for informative phenomenology to know you've reached your destination. Whereas, if some parts of problem solving occur below awareness, then the 'how' or 'why' of the solution may not be directly accessible. It is as if you've landed on a new floor with no memory of how you got there. In such cases, the phenomenology of insight is helpful for informing you that—despite not knowing how you arrived at the new location—you're on the right floor.

The above conception can explain why insight moments are problem-general rather than problem-specific, and why certain kinds of problems elicit more insight moments than others. The more likely the problem is to draw on implicit knowledge structures and processing below awareness, the more likely it is to be accompanied by an insight moment. As long as this principle is met, then there are potentially myriad cognitions that can precede an insight, which is why an idea for a new painting, a line of poetry, a way to resolve a

dispute, or a solution to an engineering problem, can all appear in mind in a sudden moment of insight. We stress this point because, if true, it is a crucial step in our understanding of insight in general: There is no single problem solving process that leads to insight—the experience may not be a consequence of a specific set of cognitions that take place in solving the problem. Instead, the experience of insight may be a signal that the work done behind the scenes has reached a conclusion that is likely to work given what is known. It is a kind of inference about the validity of the idea rather than a consequence of arriving at a solution. We now consider a trade off where the Eureka heuristic can lead to a persuasive but false sense that a true solution has been found.

The Insight Fallacy

The mathematician and Nobel laureate John Nash was asked why he believed that he was being recruited by aliens to save the world. His response powerfully illustrates the recursive danger of the Eureka heuristic. He said that: "...the ideas I had about supernatural beings came to me the same way that my mathematical ideas did. So I took them seriously" (Nasar, 1998). Here, John Nash may have committed what we term the insight fallacy. He has concluded that an idea is true solely because it occurred to him with certain phenomenology, in this case the same phenomenology as his previous mathematical discoveries. One of the benefits in defining heuristics and understanding the shortcuts we use is that we may also uncover the circumstances where they fall short. For example, the representativeness heuristic is associated with errors of base-rate neglect (Tversky & Kahneman, 1974). The availability heuristic can be lead astray by sampling biases, and anchoring and adjustment heuristics can be affected by incidental and irrelevant information (Tversky & Kahneman, 1973; 1975). What, if any, are the trade offs that occur as a consequence of using insight phenomenology to appraise the quality of our ideas?

Fear is an adaptive signal of a dangerous or challenging situation, but is also sometimes unwarranted or irrational, and in severe cases, debilitating. We can see that often fear ought to be overcome, for example, so that we can fly in a plane, swim in the ocean, or speak in public. Phobias can cause overwhelming and unjustified feelings of fear. The same is not so obvious for feelings that accompany our ideas and beliefs when they arise—our insight moments. The insight moment, like fear, may be a helpful signal that perhaps we've discovered something true and important (Laukkonen et al., 2018). However, if there is overwhelming contradictory evidence, or one is suffering from mental illness (John Nash was diagnosed with schizophrenia in 1959), then it is likely that no matter how intense our revelation, the contents of our idea will remain untrue. Just as a person might experience a profound fear of elevators, the intensity of the fear does not make the elevator dangerous. Likewise, if one is suffering from delusions, or they have been mislead with false information, then no matter how explosive the insight moment, the idea is no more likely to be correct. This is a particular concern given the powerful effect that the insight moment has on judged truth and confidence—despite the fact that we may not know why we are suddenly so confident (Danek & Wiley, 2017; Hedne et al., 2016). It is hard to predict how widespread the impact is, but consider how many contradictory ideas there are in the world and how many of the ideas that we hold most dearly were—at least subjectively—our own insights. One conclusion from this line of thinking is that we ought to be aware that the phenomenology accompanying our ideas is predictive, but also highly fallible (Danek & Wiley, 2017). Danek and Wiley (2017) found that 37% of incorrect solutions to magic tricks were reported as insight moments. The proportion of false insights in everyday life may be different—since magic tricks are a domain of negligible experience for most people—but even a fraction of 37% is alarmingly prevalent.

Since false insights do occur, then to commit the insight fallacy is to believe that an idea is true—rather than probably true—simply because the solution was accompanied by an insight experience. Hedne et al. (2016) showed that when an insight moment occurs, subjects are less likely to accept an alternative solution to the problem, and are more likely to stick with solutions that are similar to their insight. It may be that insight solutions are particularly hard to revise since the underlying process is opaque, because it is impossible to argue against reasoning that is unknown to the problem solver. The Eureka heuristic is certainly functional, but when an error does occur, the consequences can be dire. Not only are insight moments potentially incorrigible, they also promote inspiration, and provide a drive towards action (Danek & Wiley, 2017). Relative to an incorrect-but-analytic solution to a problem, when a false insight occurs, it may be more difficult to change the person's mind and to prevent them from behaving as if the solution were true.

One important implication of the insight fallacy is that humans ought to be aware that the feeling of insight is fallible and that in certain situations, it is advantageous to actively doubt light bulb moments and search for support beyond phenomenology. There are many promising avenues of research here, which are discussed further in the final section below. We propose that an important question for future research is not, "what are the variables that lead to *more* insight moments?" Instead we ought to be asking, "what are the variables that lead to *accurate* insight moments?" For example, it is likely that some psychological disorders are particularly prone to false insights. A person suffering from schizophrenia or any illness where delusions are present, may—like John Nash—be interpreting their phenomenology to conclude that an idea is true. If the functioning of this particular cognitive process is compromised, then the phenomenology experienced is no longer informative, and instead may lead to a strong conviction to ideas that are otherwise unjustified.

It may also be that some domains of knowledge are particularly prone to false insights. In the case of complex belief systems of cults, certain conspiracy theories, or superstitious beliefs, one can be an 'expert' in a domain where the knowledge structures that fuel intense Eureka moments are fictitious to begin with. The validity of the insight experience—if it is cuing a consistency with existing knowledge structures—is in direct relationship with the quality of those knowledge structures and thus the information that underlies them. Depending on unique life experiences and exposure to different cultural myths and the sheer abundance of the (often contradictory) information available, it is not surprising that there are revelations of almost every imaginable sort. Where knowledge structures are either biased or untrue, then the insight moment may in fact serve to further reinforce and motivate false beliefs. In the case of complex belief systems of cults, certain conspiracy theories, or superstitious beliefs, one can be an 'expert' in a domain where the knowledge structures that fuel intense Eureka moments are fictitious to begin with. Here, the phenomenology that accompanies new ideas may be altogether unreliable and act only to recursively increase the persistence of these worldviews (see Figure 3).

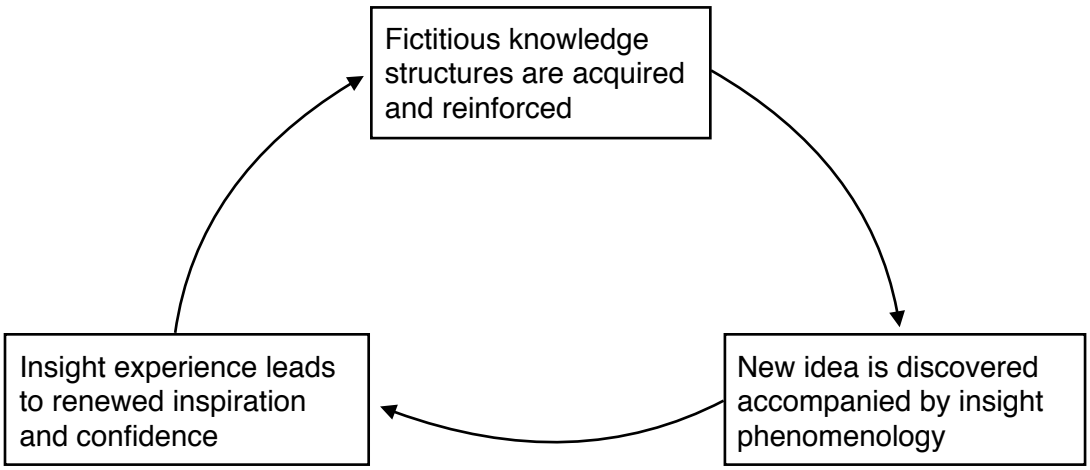

Figure 3. The recursive nature of insight in the formation of complex beliefs.

Discussion

The word Eureka originates in Ancient Greece from the word εὕρηκα (heúrēka), and before that from heuriskein, which means “I find.” Heuriskein is also the ancient origin of the word *heuristic*, which refers to shortcuts that help humans to solve problems. The shared origin of the two words *Eureka* and *heuristic* may point to a forgotten wisdom about the nature of the insight experience, that humans use the feeling of Eureka as a heuristic to evaluate the quality of their own ideas.

There is evidence that insight phenomenology plays an adaptive role in decision-making and problem solving, and we have proposed what that role may be. The heuristic view requires a restructuring of the way we think about insight moments, as a result of a specific problem solving process, to an appraisal of an idea or solution. Human experience is filled with rich phenomenology, bodily sensations, and emotions, that guide our decisions and help us to intuitively navigate complexity and uncertainty. It would be at odds with the greater body of psychological research if the ability to feel was important in most other domains of judgment and decision making, but not with regard to our own ideas and solutions to problems (Kahneman & Beatty, 1973; Kahneman, 2011; Slovic et al., 2007; Schwarz, 2010).

Since much of complex, associative, problem solving occurs below awareness, it is perhaps unnecessary—and certainly inefficient—to review all of the reasoning and information underlying every idea that comes to mind. Therefore, when it comes to solving problems, knowing ‘why we know’ is not as important as simply ‘knowing that we know’, especially when time is of the essence. To this end, insight phenomenology may serve as a fast and frugal means of signalling implicit support for a solution, employed heuristically during states of metacognitive uncertainty about problem solving processes.

The Eureka heuristic helps us understand recent data, and speaks to a number of long-standing debates in the literature. First, the Eureka heuristic can explain why insight moments occur in such a wide array of problems. Any problem can lead to an insight moment provided that some crucial steps in the solution process occurred below awareness. Traditional insight problems, remote associate problems, matchstick arithmetic, anagrams, magic tricks, and likely many others, reliably lead to insight moments in laboratories because they are easily amenable to unconscious processing. Second, the intensity of the insight experience is determined by the extent to which implicit knowledge is consistent with the new solution, which also explains the positive relationship between insight intensity and accuracy (Laukkonen et al., 2018). Third, false insights occur when an idea is consistent with one's implicit knowledge, but inconsistent with facts (Danek & Wiley, 2017). And lastly, the strong insight accuracy relationship is constrained to problems that involve unconscious processing because it is precisely in these domains where the insight experience contains information above and beyond one's deliberate reasoning.

One foreseeable criticism of the Eureka heuristic are cases of restructuring or representational change (Ohlsson, 1984). Insight problems used in the laboratory often lead to an incorrect interpretation, which leads to a mental state where prior knowledge is at odds with the solution. Consider the following example:

> *Mr. Hardy was washing windows on a high-rise office building when he slipped and fell off a sixty-foot ladder onto the concrete sidewalk below. Incredibly, he did not injure himself in any way. How is this possible?*

Most people initially assume that Mr. Hardy is standing at the top of the ladder, despite the fact that the problem does not declare it. The problem solver is now constrained by what they believe they know based upon an implicit interpretation. The problem can trigger an insight

solution only when the problem is implicitly restructured and the problem solver realises that Mr. Hardy is on the bottom of the ladder (Ohlsson, 1984). Then one might ask: How is it possible that insight signals consistency with prior knowledge, since prior knowledge was the very barrier to solving the problem?

While existing knowledge is initially at odds with the true solution, it is precisely when the unconscious restructuring occurs that the problem solvers knowledge suddenly *becomes* consistent with the solution. With the incorrect representation of the problem, no insight moment occurs because no solution feels coherent with one's implicit interpretation. However, when one's assumptions change—i.e., the state of one's implicit knowledge and beliefs change—then an insight moment occurs as the solution is suddenly consistent with what one knows about the problem.

There is, however, a dark side to the Eureka heuristic. There is evidence that insight moments are difficult to revise (Hedne et al., 2016), and are more easily recalled (Danek et al., 2013). The insight moment may be highly functional most of the time, but when it signals a false solution, then the implications are profound. The individual may be left with a powerful sense of certainty, and also the drive and inspiration to act according to an incorrect solution. In ill-defined domains, where one has very little experience, or domains that are fictional by nature—insight moments may simply provide a kind of recursive illusion of progress. In such circumstances, the insight moment may reinforce solutions and ideas that are false, and in some cases inspire further work in the wrong problem space.

Future Directions

Throughout this paper, we have made a number of predictions and highlighted several areas for future research. This may be seen as both a strength and a weakness of the account; a number of possible streams of research are identified, but a number of hypotheses remain open. The promise of novel experiments, and the breadth of the theories and empirical work on which our predictions are founded, we believe sufficiently outweighs the risks of falsification (the possibility of which is itself a strength). Below we reiterate some key predictions already made and provide a few more novel hypotheses.

In the spirit of heuristics and biases, one particularly promising direction for empirical work will be in identifying the key variables that predict the accuracy of insight moments, thereby identifying potential biases arising from the Eureka heuristic. Many studies have aimed to measure and manipulate the ability to solve insight problems, or increase the incidence of insight experiences (e.g., Jarosz, Colfesh, & Wiley, 2012; Ostafin & Kassman, 2012; Steidle & Werth, 2013; Thomas & Lieras, 2009; Weller et al., 2011). However, increasing the incidence of insights overall may also lead to more false insights. In the context of everyday life, false insights may be at least as unhelpful as accurate insights are helpful.

Another key prediction is that, since people seem to rely on phenomenology rather than deliberate reasoning to evaluate the quality of an idea, then verbalising the reasoning behind an insight solution may lead to confabulation. For example, it is conceivable that a problem may be correctly solved in an insight experience, but because the true reasoning behind the insight is implicit, then the explicit reasoning may turn out to be false. Just as humans may have mistaken interpretations about the reasons for their behaviour, they may have mistaken interpretations about the reasons for their own ideas and solutions to

problems. Another possibility is that insight phenomenology may be employed to induce 'truth misattribution,' in a similar fashion to discovery misattribution (Dougal and Schooler, 2007). For instance, if an insight experience can be made to occur at the same time as an unrelated proposition (e.g., "lithium is the lightest of all metals," Reber & Unkelbach, 2010), then the insight experience may be falsely attributed to—or confused with—the truth of the proposition. More broadly, a bias towards perceived veracity could be induced in any circumstance where insights occur together with the stimulus.

Manipulations of insight phenomenology should also lead to changes in confidence and truth judgments regarding associated ideas. One such candidate is positive affect, as noted by Topolinski and Reber (2010). Since positive affect is a key dimension of most insight experiences (Danek & Wiley, 2017), then participants who experience positive affect may report more frequent and perhaps more intense reports of insight during problem solving. There is indeed some evidence that positive affect increases reports of insight (Subramaniam et al., 2009). If the positive affect is not relevant to the informational content of the insight moment—for example if it is artificially induced alongside problem solving—then we would expect more false insights. Other practical candidates for confounding the insight experience include arousal, fluency, suddenness, and surprise.

Another interesting question is whether it is possible to shift individual criteria about what counts as a true insight, and whether such individual differences exist. Those who score high on the 'need for cognition,' for instance, may adopt a more conservative criterion and therefore experience fewer false insights, whereas those with a low score may adopt a more liberal criterion (Cacioppo & Petty, 1982). Similarly, people higher in mindfulness may be better able to monitor their own phenomenology, and thereby less likely to mistakenly believe that an insight occurred (a hypothesis we have found preliminary evidence for). Another

possibility is that some intoxicants or psychoactive substances may jeopardise how informative insight experiences are, and thereby influence the rate of false insights. In any case, it is possible that problem solving performance overall may be similar among different groups, but the incidence of true versus false insight moments may differ significantly. Finally, it may turn out that particular cognitive dysfunctions are strong predictors of false insight experiences. If this turns out to be true, then it may also be possible to use the incidence of false insights during problem solving tasks as a diagnostic tool.

Concluding remarks.

A functional, heuristic view of insight experiences is likely to help move the field in many new and productive directions. In particular, questions regarding the behavioral consequences of insight—or biases associated with judgments that rely on insight—become more pertinent than questions regarding problem solving processes, which have held center stage for the past century. We are optimistic that the Eureka heuristic account of insight fits the existing data, provides an explanation regarding the functional aspects of insight phenomenology, and can explain why that phenomenology has predictive power. This account helps to explain why insight moments occur across a range of different problems, and illuminates the pitfalls of relying on phenomenology to guide problem solving in some circumstances.

References

Anderson, N.H., 1981. *Foundations of information integration theory.* Academic, New York.

Bowden, E. M. (1997). The effect of reportable and unreportable hints on anagram solution and the aha! experience. *Consciousness and Cognition, 6*(4), 545-573 .

Bowden, E., & Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods, 42*(1), 87-99.

Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *The Behavioral and Brain Sciences, 32*.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science,* 264(5162), 1102-1105.

Danek, A., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). Working Wonders? Investigating insight with magic tricks. *Cognition, 130*(2), 174-185.

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

Dougal, S., & Schooler, J. W. (2007). Discovery misattribution: when solving is confused with remembering. *Journal of Experimental Psychology: General, 136*(4), 577. Chicago

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist, 49*(8), 725.

Fischhoff, B., Slovic, P., Lichtenstein, S., Reid, S., Coombs, B., 1978. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences 9*, 127–152.

[fsgm1]. (2011, Jan 31). *The great moment at which Andrew Wiles solved Fermat's last theorem, in 1994.* Retrieved from https://www.youtube.com/watch?v=SccDUpIPXM0

Gick, M.L., & Lockhart, R.S. (1995). Cognitive and affective components of insight. In R.J. Sternberg & J.E. Davidson (Eds.), *The Nature of insight.* (pp. 197–228). Cambridge, MA: MIT Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62,* 451-482.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review, 109*(1), 75.

Hattori, M., Sloman, S. A., & Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin & Review, 20*(4), 790-797.

Hedne, M. R., Norman, E., & Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. *Frontiers in Psychology, 7.*

Jarosz, A. F., Colflesh, G. J., & Wiley, J. (2012). Uncorking the muse: Alcohol intoxication facilitates creative problem solving. *Consciousness and Cognition, 21(1), 487-493.*

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2004). Facing changes: Choice blindness and facial attractiveness. *International Journal of Psychology, 39*(5-6), 287-287.

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310*(5745), 116-119.

Johansson, P., Hall, L., Sikstrom, S., Tarning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition, 15*(4), 673-692.

Kahneman, D. (2015). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Klein, G. & Jarosz, A. (2011). A Naturalistic Study of Insight. *Journal of Cognitive Engineering and Decision Making,* 5(4), 335-351.

Kounios, J. & Beeman, M. (2014). The Cognitive Neuroscience of Insight. *Annual Review of Psychology, 65*(1), 71-93.

Landrum, R. E. (1990). Maier's (1931) two-string problem revisited: evidence for spontaneous transfer?. *Psychological Reports, 67*(3), 1079-1088.

Latané, B., & Darley, J. M. (1970). The unresponsive bystander: Why doesn't he help?. *Appleton-Century-Crofts.*

Laukkonen, R. & Tangen, J. (2017). Can observing a Necker cube make you more insightful?. *Consciousness and Cognition, 48,* 198-211.

Laukkonen, R., Ingledew, D., Schooler, J., & Tangen, J. M. (2018, March 15). *The phenomenology of truth: The insight experience as a heuristic in contexts of uncertainty.* Retrieved from psyarxiv.com/9w56m

Loewenstein, G.F., Weber, E.U., Hsee, C.K., Welch, E.S., 2001. Risk as feelings. *Psychological Bulletin 127*, 267–28

Maier, N. R. (1931). Reasoning and learning. *Psychological review, 38*(4), 332.

Mellers, B.A., Richards, V., Birnbaum, J.H., (1992). Distributional theories of impression formation. *Organizational Behavior and Human Decision Processes 5*1, 313–343.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15*(3), 238-246.

Nasar, S. (1998). *A Beautiful Mind: A Biography of John Forbes Nash, Winner of the Nobel Prize in Economics, 1994*. Simon and Schuster.

Nisbett, R. E., & Schachter, S. (1966). Cognitive manipulation of pain. *Journal of Experimental Social Psychology, 2*(3), 227-236.

Nisbett, R. & Wilson, T. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review 84*:231–95.

Nisbett, R. E. (2015). *Mindware: Tools for smart thinking*. Farrar, Straus and Giroux.

Ohlsson, S. (1984). Restructuring revisited. *Scandinavian Journal Of Psychology, 25*(2), 117-129.

Ostafin, B. D., & Kassman, K. T. (2012). Stepping out of history: Mindfulness improves insight problem solving. *Consciousness and Cognition, 21*(2), 1031-1036.

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. *The Journal of Creative Behavior, 0*, 1-18

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*(1), 45-48.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*, 338–342.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology, 1*(4), 563-581.

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Thinking & Reasoning,* 1-18.

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. Smith, T. Ward, & R. Finke (Eds.), *The Creative Cognition Approach* (pp. 97–133). Cambridge, Mass.: MIT Press.

Schooler, J. W. (2001). Discovering memories of abuse in the light of meta-awareness. *Journal of Aggression, Maltreatment & Trauma, 4*(2), 105-136.

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6*(8), 339-344.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*(3), 610.

Schwarz, N. (2011). *Feelings-as-information theory. Handbook of theories of social psychology, 1*, 289-308.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.

Snyder, A., Mitchell, J., Ellwood, S., Yates, A., & Pallier, G. (2004). Nonconscious idea generation. *Psychological Reports, 94,* 1325–1330.

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy, 2*(1), 85-95.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research, 177*(3), 1333-1352.

Steidle, A., & Werth, L. (2013). Freedom from constraints: Darkness and dim illumination promote creativity. *Journal of Environmental Psychology, 35*, 67-80.

Subramaniam, K., Kounios, J., Parrish, T., & Jung-Beeman, M. (2009). A Brain Mechanism for Facilitation of Insight by Positive Affect. *Journal Of Cognitive Neuroscience, 21*(3), 415-432.

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

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence, and of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General, 138,* 39–63.

Topolinski, S. & Reber, R. (2010). Gaining Insight Into the "Aha" Experience. *Current Directions In Psychological Science, 19*(6), 402-405.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology, 5*(2), 207-232.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. *In Utility, probability, and human decision making* (pp. 141-162). Springer Netherlands.

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

Wegner, D. & Wheatley, T. (1999) Apparent mental causation: Sources of the experience of the will. *American Psychologist 54*:480–91.

Wegner, D. (2002) *The illusion of conscious will.* MIT Press.

Weller, A., Villejoubert, G., & Vallée-Tourangeau, F. (2011). Interactive insight problem solving. *Thinking & Reasoning, 17*(4), 424-439.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29*(6), 716-732.

Whittlesea, B. W., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98*(2), 141-165.

Whittlesea, B. W., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129*(1), 84.

Winkielman, P., Zajonc, R.B., Schwarz, N., 1997. Subliminal affective priming resists attributional interventions. *Cognition and Emotion 11* (4), 433–465.

Winkielman, P., & Cacioppo, J.T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation leads to positive affect. *Journal of Personality and Social Psychology, 81*, 989–1000.