The ring of truth: Irrelevant insights make worldviews seem true

Ruben Laukkonen (ruben.laukkonen@gmail.com)  
Vrije Universiteit Amsterdam

Benjamin Kaveladze  
University of California, Irvine, USA

John Protzko  
University of California, Santa Barbara  https://orcid.org/0000-0001-5710-8635

Jason Tangen  
The University of Queensland

William von Hippel  
University of Queensland

Jonathan Schooler

Article

Keywords:

DOI: https://doi.org/10.21203/rs.3.rs-547655/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Our basic beliefs about reality can be impossible to prove and yet we can feel a strong intuitive conviction for them, as exemplified by insights that imbue an idea with immediate certainty. Here we presented participants with worldviews such as “people’s core qualities are fixed”, and simultaneously elicited an aha moment. In the first experiment (N = 3000), which included a direct replication, participants rated worldview beliefs as truer when they solved anagrams and experienced aha moments. A second experiment (N = 1,005) showed that the worldview statement and the aha moment must be perceived simultaneously for the insight misattribution effect to occur. These results demonstrate that artificially induced aha moments can make worldviews seem truer, possibly because humans rely on feelings of insight to appraise an idea's veracity. Feelings of insight are therefore not epiphenomenal and should be investigated for their effects on decisions, beliefs, and delusions.

Introduction

The philosopher Rob Sips (2019) said that aha moments played an essential role in the deterioration of his psychosis. He describes experiencing an “accelerating stream” (p2) of aha moments that revealed the world from so many different perspectives that his worldview simply could not withstand the assault: “This process, in my experiences, was wrecking what I considered to be my “personal worldview”... [the insights] undermined or “derealized” how I looked at things before.” (p3). The idea that aha moments—a sudden feeling of pleasure and certainty that accompanies a new idea—can arise from a change in perspective goes back almost a century (Duncker, 1945; Maier, 1930) and is the basis of how psychologists elicit insight in the laboratory (Ohlsson, 1984). Like the switching perspectives in the duck-rabbit illusion, aha moments mark a novel discovery in which pre-reflective assumptions change and information is seen in a new light (Schooler & Melcher, 1995). The experience of Rob Sips (2019) exemplifies the dramatic impact of this process on higher-order worldviews. For Sips, the uncontrolled cascade of aha moments marked the deconstruction of his beliefs, leading to an unstable grasp on reality—like a multidimensional duck-rabbit illusion that will not stop shifting.

Sips’ (2019) account is dramatic, but it is possible that aha moments are simply epiphenomenal, marking but not causing changes in beliefs, much like the steam-whistle of an engine. Under this view, the aha moment is a feeling that correlates with the discovery of a new perspective or solution but has no impact on decision-making or the selection of new ideas (Klein & Jarosz, 2011). Alternatively, as Sip’s experience suggests, the aha moment may be causally potent. For example, aha moments might provide feedback to the conscious agent about whether an idea is likely to be a good one. Under this view, the aha experience convinces the agent that the new perspective is true. Because the processes that precede aha moments can be pre-reflective or implicit (Bowden, 1997; Grant & Spivey, 2003; Hattori et al., 2013; Laukkonen & Tangen, 2017; 2018; Laukkonen et al., 2021; Maier, 1931; Salvi et al., 2015; Salvi & Bowden, 2020; Schunn & Dunbar, 1996; Sio & Ormerod, 2009), aha moments might provide information that is not otherwise accessible. Thus, the aha moment could plausibly provide information about whether an idea
can be trusted in the absence of access to the processes that produced it, much like hunger or fear can signal something important about the state of one's inner or outer world (Damasio, 1996; Schwarz, 2012).

Consistent with this possibility, preliminary evidence suggests that aha experiences provide important information. For instance, aha moments correspond to more accurate solutions to problems (Danek & Wiley, 2017; Hedne et al., 2016; Laukkanen et al., 2021; Salvi et al., 2016; Threadgold et al., 2018; Webb et al., 2016). The correspondence between the magnitude of the aha feeling and accuracy has also been captured using a measure of grip strength (Laukkanen et al., 2021); the harder the participants (unintentionally) squeezed the device during the spontaneous aha moment, the more likely it was to be correct. Although these findings do not demonstrate that aha moments are causally agentive, they suggest that aha moments carry valuable information that could be useful for decision-making. There is also more direct evidence that aha moments can affect decisions. For example, aha moments that occur when solving anagrams lead to a kind of false memory, wherein participants report having already seen the word even if they had not (Dougal & Schooler, 2007). Using a similar paradigm, another study showed that irrelevant aha moments make mundane facts more believable (Laukkanen et al., 2020).

In the experiments that follow we test, replicate, and extend the following hypothesis: That the seemingly trivial aha moment elicited by solving anagrams can increase the perceived veracity of important beliefs that serve as the basis of people's worldviews. Because feelings of insight carry useful information about the quality of the associated ideas, when an aha moment is experienced it should lead to a truer appraisal of the worldview that accompanies it. Metaphorically, the aha moment “lights up” more content than it should, making the temporally coincident but unrelated belief seem true. According to this possibility, Rob Sips’ slide into psychosis may not have been driven by accelerating insights about the state of the world. Rather the aha moments themselves may have contributed to the unraveling of his belief system due to the new perspectives that they happened to reinforce. According to this possibility, the aha experience may act as a heuristic for appraising the quality of ideas appearing in mind, which we have previously termed the Eureka Heuristic (Laukkanen et al., 2018).

**Experiment 1: Demonstration And Replication**

This experiment was approved by the University of California, Santa Barbara, Human Subjects Committee, in accordance with the Declaration of Helsinki.

**Methods**

**Design & materials**

The experiment had two within subject variables: 2 (Problem: solved or unsolved) × 2 (Aha Experience: present or absent), and one between-subjects factor (Anagrams: present or absent). The dependent measure was truth judgments on a 12-point scale ranging from 1 (definitely false) to 12 (definitely true).
We created 15 worldview claims, none of which were objectively demonstrable as true or false. Each claim was constructed such that the last word of it was key to its meaning (see procedure for an example). The worldview claims were derived conceptually from ‘The Psychology of Worldviews’ (Koltko-Rivera, 2004). We also created 15 anagrams derived from the last word of each claim (see Appendix A for stimuli). Keywords were re-organised into anagrams using a random scramble function, then iteratively pilot tested and adjusted manually until they were neither too difficult nor too easy (approximating 50% solution rates).

**Participants and procedure**

This study had two samples of 1,500 participants recruited by *Critical Mix* to match the demographics of the U.S. We used the first sample of 1,500 to test our hypotheses, and then the second 1,500 to assess the replicability of the findings. We determined that 1,500 participants would provide sufficient power to detect a Cohen’s d effect size of .4, based on Laukkonen et al. (2020). All participants were randomly assigned to either the *Anagram* or *No Anagram* (control) condition via written instructions. The procedure for the *Anagram* condition is illustrated in Fig. 1 below. In the *No Anagram* condition, participants were simply presented with the complete propositions (e.g., “It is useless to pursue justice”) and judged how true they were.

**Results**

In our analyses, we rely primarily on two statistical approaches. For between-subjects comparisons across conditions, we use Welch’s *t*-tests. For within-subjects analyses, we use multilevel regression models, which account for the hierarchical nature of the data with random intercepts and random slopes for participants and trial numbers (Westfall et al., 2014). For the multilevel regression models with one binary fixed effect, we present the Cohen’s *d* of the fixed effect, which we calculate according to Westfall et al. (2014). These models assess the influence of the predictors *aha moment* (Y/N) and *correct anagram solution* (Y/N) with truth scores as the DV, as well as the predictor *aha moment* with correct solutions as the DV. We report statistics in accordance with similar previous work (e.g., Ding et al., 2021). Between-subjects analyses were conducted using the *t*-test function within the stats package in R (R Core Team, 2013). All within-subjects analyses were conducted using the lmer function within the lme4 (Bates et al., 2007) package in R, and effect sizes were calculated using the r.squaredGLMM function within the MuMin package (Bartoń, 2019). Our data and analysis scripts are available on the OSF (demonstration data: https://osf.io/2duq5, replication data: https://osf.io/rzg3c, complete analyses: https://osf.io/wycmg).
Demonstration: First 1,500 Participants

After excluding 247 participants from analysis for failing to solve any anagrams ($N = 82$), solving all anagrams ($N = 4$), experiencing no aha moments ($N = 178$), or experiencing aha moments on all trials ($N = 23$), 1,250 participants were included in the analyses. 443 participants were in the Anagram condition and 807 were in the No Anagram condition (note: we address the differential dropout rate in Experiment 2). On average, participants correctly solved the anagrams 37% of the time ($SD = 21$%), and the anagrams elicited aha moments 36% of the time ($SD = 22$%). Using a multilevel regression model with aha moments as a fixed effect and participants as a random effect, we found that anagrams that elicited aha moments were more likely to be solved correctly (23% solved, $SD = 18$%) than anagrams that did not elicit aha moments (14% solved, $SD = 18$%), $b = .41$, $t = 16.10$, $p < .001$, $d = .72$.

Truth judgments: Anagrams versus no anagrams (between): As predicted, a between-subjects Welch’s t-test revealed that participants’ average truth scores in the Anagram condition were higher ($M = 6.62$; $SD = 2.09$) than participants’ average truth scores in the No-Anagram condition ($M = 5.75$; $SD = 2.05$), $t(781) = 7.54$, $p < .001$, $d = .46$. Overall, the presence of the anagram—including both solved and unsolved anagrams—increased truth judgments regarding the worldview claims.

Truth judgments: Solutions and aha moments (within): We used a multilevel regression model to test our prediction that claims associated with solved anagrams would be rated as more likely to be true than claims associated with unsolved anagrams, including solving as a fixed effect and participants as a random effect. As predicted, solved anagrams resulted in higher truth ratings ($M = 7.06$, $SD = 2.70$) than unsolved anagrams ($M = 6.45$, $SD = 2.15$), $b = .60$, $t = 4.97$, $p < .001$, $d = .16$. We also used a multilevel regression model to test our prediction that claims would be rated as more likely to be true if they were accompanied by aha moments while solving the anagram, including aha moments as a fixed effect and participants as a random effect. As predicted, participants provided higher truth ratings on trials where they reported experiencing an aha moment ($M = 7.21$, $SD = 2.73$) than on trials where they did not experience an aha moment ($M = 6.35$, $SD = 2.26$), $b = .89$, $t = 6.85$, $p < .001$, $d = .23$. Finally, we examined aha moments among the subset of correctly solved anagrams. Correctly solved anagrams accompanied by aha moments had higher truth ratings ($M = 7.28$, $SD = 2.92$) than correctly solved anagrams without aha moments ($M = 6.29$, $SD = 2.90$), $b = .89$, $t = 5.13$, $p < .001$, $d = .23$.

Replication: Second 1,500 participants

After excluding 261 participants from analysis for failing to solve any anagrams ($N = 123$), solving all anagrams ($N = 1$), experiencing no aha moments ($N = 155$), or experiencing aha moments on all trials ($N = 47$), 1239 participants were included in analyses. 434 participants were in the Anagram condition and
805 were in the No Anagram condition. On average, participants correctly solved the anagrams 36% of the time (SD = 21%), and the anagrams elicited aha moments 37% of the time (SD = 23%). Using a multilevel regression model with aha moments as a fixed effect and participants as a random effect, we found that anagrams that elicited aha moments were more likely to be correctly solved (M = 21%, SD = 17%) than anagrams that did not elicit aha moments (M = 15%, SD = 17%), b = .35, t = 15.25, p < .001, d = .62.

**Truth judgments: Anagrams versus no anagrams:** As predicted and consistent with the first sample, participants’ average truth scores in the Anagram condition were higher (M = 6.81; SD = 2.09) than participants’ average truth scores in the No Anagram condition (M = 6.22; SD = 1.92), t(822) = 4.84, p < .001, d = .29. As in the first sample, the presence of the anagram—including both solved and unsolved trials—increased truth judgments.

**Truth judgments: Solving and aha moments:** We used a multilevel regression model to test our prediction that claims associated with solved anagrams would be rated as more likely to be true than claims associated with unsolved anagram, including solving as a fixed effect and participants as a random effect. As predicted, solved anagrams resulted in higher truth ratings (M = 7.09, SD = 2.70) than unsolved anagrams (M = 6.69, SD = 2.24), b = .49, t = 4.51, p < .001, d = .13. We also used a multilevel regression model to test our prediction that claims would be rated as more likely to be true if they were accompanied by aha moments while solving the anagram, including aha moments as a fixed effect and participants as a random effect. As predicted, participants provided higher truth ratings on trials where they reported experiencing an aha moment (M = 7.34, SD = 2.71) than on trials where they did not experience an aha moment (M = 6.52, SD = 2.23), b = .91, t = 7.58, p < .001, d = .25. Finally, we conducted the aha moment analysis among the subset of correctly solved anagrams. Correctly solved anagrams accompanied by aha moments had higher truth ratings (M = 7.29, SD = 2.96) than correctly solved anagrams without aha moments (M = 6.45, SD = 2.74), b = .82, t = 4.53, p < .001, d = .23.

**Confirmation + Replication**

In Fig. 2, we illustrate the combined results of the confirmation (N = 1,250) and the replication (N = 1,239). For each of the illustrated comparisons: Anagram (present vs. absent), Solving (yes vs. no), and Aha (present vs. absent), and aha for solved anagrams (present vs. absent).

**Experiment 2: Aha Misattribution With Delay**

This experiment was approved by the University of California, Santa Barbara, Human Subjects Committee, in accordance with the Declaration of Helsinki.
Methods

Design & materials

In this study we test the assumption that the aha experience needs to be temporally coincident with the claim, by introducing a “delay condition” with a 10-second interval between solving anagrams and being shown claims. Additionally, we address two limitations of the first experiment. First, to reduce differential dropout across conditions, we standardize the difficulty and completion time of each condition. Second, we introduce two new conditions that shift the order of trial components to rule out potential order effects with regard to anagram solving.

The dependent measure was again truth judgments on a 12-point scale ranging from 1 (definitely false) to 12 (definitely true) regarding the 15 claims from Experiment 1 (see Appendix A). To standardize the difficulty, we provided hints when participants attempted to solve the anagrams. Raw data and materials can be found on the OSF: [https://osf.io/vym8r](https://osf.io/vym8r).

Participants and procedure

For this study we had a sample of 1,564 participants recruited by Critical Mix to match the demographics of the U.S. Using the pwr function in R, we determined that 99 participants in each of the two key groups of interest would provide sufficient power (.8) to detect a Cohen's $d$ effect size of .4 for the main analysis (Champely, 2020). Thus, factoring potential dropouts, 1,564 participants randomly assigned to one of four conditions (illustrated in Table 1) should provide more than sufficient power. Participants were provided written instructions and we included hints alongside anagrams to achieve comparable solving rates for all the conditions. A ten-second delay was provided at the end of each trial.

**Table 1.** Four independent conditions in Experiment 2
**Anagram Normal** $N = 149$

1. Presentation of an incomplete proposition and concurrently resolving an anagram that completes the proposition (with the help of a hint) within 20 seconds or presentation of the solution after 20 seconds if no response is submitted

2. A truth judgment about the completed proposition

3. Reporting whether an aha moment occurred

**Anagram Delay $N = 157$**

1. Resolving an anagram with the help of a hint within 20 seconds or presentation of the solution after 20 seconds if no response is submitted

2. A 10-second "delay" period

3. A truth judgment about the completed proposition

4. Reporting whether an aha moment occurred

**Anagram After Truth $N = 132$**

1. Making a truth judgment about a completed claim

2. Unscrambling the keyword from a different claim with a hint

3. Reporting whether an aha moment occurred

**Anagrams After Everything $N = 121$**

1. Providing truth ratings for all claims

2. Unscrambling all anagrams with the help of a hint, reporting aha experiences, and a 20-second delay between anagrams

---

**Aim:** This is the experimental condition from Experiment 1

**Aim:** This condition tests whether the aha moment needs to be temporally coincident with the claim

**Aim:** This condition establishes baseline truth ratings for each claim without any solving effects (random anagrams are solved after participants make a truth judgment)

**Aim:** This condition establishes true baseline truth ratings, where all anagrams are solved after all truth ratings are completed

---

**Results**

We generally used the same analytic methods as outlined in experiment 1, except that instead of using Welch’s $t$-tests to examine the between-subjects effects, we used independent samples ANOVAs with truth ratings as the DV and condition as the factor. Standardizing the difficulty and completion time resulted in similar Ns in the four conditions: Anagram Normal = 149, Anagram Delay = 157, Anagram After Truth = 132, and Anagram After Everything = 121. Based on our preregistered exclusion criteria, 329 participants were excluded for failing to solve any anagrams ($N = 158$), solving all anagrams ($N = 23$), experiencing no aha moments ($N = 174$), or experiencing aha moments on all trials ($N = 38$). As a result, 676 participants were included in analyses (see Table 1 for a breakdown of the various conditions).
The mean solution rate for anagrams presented alongside hints and an unfinished worldview claim was 53% (SD = 24%). When only the hint was provided, a similar solving rate was found, 57% (SD = 24%), suggesting that the hints helped to balance solving rates. We also successfully equated aha moments across the two key conditions: Anagram Normal condition (M = 44%, SD = 26%), Anagram Delay condition (M = 42%, SD = 26%). Using a multilevel regression model with aha moments as a fixed effect and participants as a random effect, we found that anagrams that elicited aha moments (in the Anagram After Delay and Anagram Normal conditions) were more likely to be correctly solved (M = 33%, SD = 24%) than those that did not (M = 22%, SD = 24%), b = .38, p < .001, t = 13.14, d = .65.

Truth judgments: Comparison across conditions (between): Our key prediction was that truth judgments would be lower in the Anagram Delay condition than in the Anagram Normal condition. An independent samples ANOVA revealed a main effect of condition, F(3, 672) = 12.52, p < .001, η² = .053. Follow-up Tukey comparisons support our key prediction that a delay would remove the aha misattribution effect: the Anagram Normal condition elicited higher truth ratings (M = 6.94; SD = 2.16) than the Anagram Delay condition (M = 5.98; SD = 1.66), t(672) = 4.60, p < .001, d = .50. Moreover, the Anagram Normal condition elicited higher truth ratings than the Anagram After Truth condition (M = 5.97; SD = 1.61), t(672) = 5.04, p < .001, d = .51, and the Anagram After Everything condition (M = 5.90; SD = 1.72), t(672) = 5.39, p < .001, d = .53. There were no other significant effects, indicating that the order of anagram solving was redundant and that the anagram simply needed to occur at the same moment as the worldview was presented.

Truth judgments: Solving and aha moments (within): We used a multilevel regression model to test our prediction that participants in the Anagram Normal condition would rate claims associated with solved anagrams as more likely to be true than claims associated with unsolved anagram, including solving as a fixed effect and participants as a random effect. We also predicted that participants in the Anagram Normal condition would rate claims as more likely to be true if they experienced an aha moment while solving the anagram. We further predicted that both of these effects would be absent or weaker in the Anagram Delay condition.

The predicted interaction between condition (Anagram Normal vs. Delay) and solved anagrams was significant, b = .42, t = 2.06, p = .041. Inconsistent with predictions, the higher truth ratings in the presence of solved anagrams (M = 7.27, SD = 2.56) than unsolved anagrams (M = 6.67, SD = 2.40) in the Anagram Normal condition failed to reach conventional levels of significance, b = .35, t = 1.79, p = .083, d = .09. In the Anagram Delay condition, no effect of solving emerged on truth ratings (M = 6.00, SD = 1.91 vs. M = 6.11, SD = 1.96), b = -.18, t = -.87, p = .396, d = -.05.
As predicted, the interaction between condition (Anagram Normal vs. Delay) and experiencing aha moments was significant, $b = .70, t = 3.47, p < 0.001$. Simple effects analyses revealed that experiencing aha moments when solving anagrams resulted in higher truth ratings ($M = 7.54, SD = 2.62$ vs. $M = 6.60, SD = 2.19$) in the Anagram Normal condition, $b = .85, t = 5.54, p < .001, d = .23$, but not in the Anagram Delay condition ($M = 6.20, SD = 2.36$ vs. $M = 6.02, SD = 1.82$), $b = .12, t = .76, p = .447, d = .03$.

Finally, we conducted the aha moments analysis among the subset of correctly solved anagrams. Inconsistent with predictions, the interaction between condition and experiencing aha moments was not significant, $b = .33, t = 1.00, p = 0.32$. Nonetheless, simple effects analyses revealed that correctly solved anagrams accompanied by aha moments had higher truth ratings than those not accompanied by aha moments ($M = 7.45, SD = 2.88$ vs. $M = 6.73, SD = 2.65$) in the Anagram Normal condition, $b = .58, t = 2.65, p = .009, d = .16$ (replicating Experiment 1), but not in the Anagram Delay condition ($M = 6.17, SD = 2.47$ vs. $M = 5.93, SD = 2.45$), $b = .23, t = .92, p = .361, d = .06$. The key findings are illustrated in Fig. 3.

**Discussion**

The present study tested whether incidental aha experiences could influence the perceived veracity of different worldviews. In the first experiment, we found that participants rated worldview statements as truer when they had just attempted to solve anagrams corresponding to those statements. We also found that successfully solving the anagram led to higher truth ratings than failing to solve them. And finally, for the subset of correctly solved anagrams, those that elicited aha moments had the highest truth ratings of all. We then directly replicated this effect. In a second experiment, we manipulated the timing of anagram solving and therefore also aha moments. Here we found that aha moments increase perceived truth only when they occur at the same time as when the worldview is presented. In short, it seems that well-timed artificially induced aha experiences can impact people's assessments of central premises about the world, giving them a ring of truth that they would not otherwise enjoy.

The fact that the feeling of insight has an impact on one's judgments is not itself surprising. There is a long list of domains where feelings influence decisions, including jury decision-making (Semmler & Brewer, 2002), risk judgments (Fischhoff et al., 1978), truth and memory judgments (Dougal & Schooler, 2007; Reber & Schwarz, 1999; Schwarz et al., 2007), and gambling and probability judgments (Loewenstein et al., 2001). But why would the aha experience influence truth ratings about something as seemingly unrelated and fundamental as worldview beliefs?

Aha experiences are characterized by an immediate sense of confidence and pleasure in the content of an idea or solution (Danek & Wiley, 2017; Webb et al., 2018). This feeling of certainty is warranted—aha moments tend to correspond to accurate solutions (Danek & Wiley, 2017; Hedne et al., 2016; Laukkonen et al., 2018; Salvi et al., 2016; Threadgold et al., 2018; Webb et al., 2016). Because aha experiences tend to be a marker of good ideas, it makes sense that humans have learned to draw on this feeling as a source of information about our beliefs (Laukkonen et al., 2020; Laukkonen et al., 2021). Our findings thus favor
the hypothesis that aha moments are not simply epiphenomenal—like the steam whistle of an engine—but have causal influence on guiding decisions about the veracity of new ideas, like the coal that fuels the engine (Dougal & Schooler, 2007; Laukkonen et al., 2020).

It is worthwhile briefly distinguishing our aha misattribution account from fluency or ease of processing effects (Reber & Schwarz, 1999; Topolinski & Reber, 2010). Under the fluency account, when an anagram is solved—regardless of aha moments—there ought to be an increase in fluency. However, within solved anagrams, the presence or absence of aha moments led to higher truth ratings, thus going above and beyond fluency alone. The fluency account would also presumably predict higher truth values when there are no anagrams to solve, since the presence of anagrams ought to lead to a more disfluent experience overall. Yet, we found the opposite in Experiment 1.

A fruitful path for future work is to investigate the effects of aha on decision-making, particularly in applied contexts. Aha moments can be incorrigible (Hedne et al., 2016), difficult to forget (Danek et al., 2013), and can promote inspiration and drive towards action (Danek & Wiley, 2017). It is also well known that humans often fail to introspect about the true causes of their feelings or actions (Brasil-Neto et al., 1992; Carruthers, 2009; Johansson et al., 2005; Nisbett & Wilson, 1977; Wegner, 2004; Wegner & Wheatley, 1999). Thus, the persuasive insights that humans experience in complex domains such as politics, religion, and relationships, may all be influenced by the recursive force of aha moments. Aha moments can mark a valuable new discovery, but if this process breaks down or is misinformed, then they may also perpetuate and entrench false beliefs. A potentially disastrous example of this mechanism in action can be seen in the QAnon phenomenon. Here an unknown individual(s) set up vague clues for the public to identify patterns across the media, political and presidential proceedings, and other current events, in order to confirm a grand conspiracy in which the former president of the United States was acting behind the scenes to stop a pedophilic cannibalistic cabal. The level of support for the movement is hard to measure, but appears remarkably high given the bizarre nature of the claims (Shanahan, 2021). QAnon provides a potential real-life example of our findings in its demonstration that the way the mind constructs ‘insights’ is fallible (Grimmer et al., 2021), and yet these insights can incite real and sometimes dangerous behaviour.

Moving forward, we encourage a research program on the underlying mechanisms and predictors of false insights—circumstances and states of mind in which this usually adaptive heuristic may break down. A paradigm has recently been developed to experimentally elicit false insights (Grimmer et al., 2021), which could yield valuable data for understanding the development of delusions in clinical populations. Consider, for example, the case of John Nash, the Nobel Laureate and mathematician. When he was asked why he believed he was being recruited by aliens to save the world. He said, "...the ideas I had about supernatural beings came to me the same way that my mathematical ideas did. So I took them seriously" (Nasar, 2001). Along with the example of Rob Sips (2019) described earlier, such anecdotes may point to the failure of an otherwise adaptive Eureka heuristic (Laukkonen et al., 2018). Under ordinary conditions, the feeling that accompanies a new idea reveals that inaccessible processes have yielded a valuable conclusion. This heuristic view of the aha moment also makes sense evolutionarily, as...
we often must decide quickly whether a new idea is good or bad; a thorough analysis is not always possible when an audience member asks a challenging question or a hungry lion is at one's heels.

References

Bartoń, K. (2019). MuMIn: Model Selection and Model Averaging Based on Information Criteria (AICc and alike) (R Package Version, 1–1, 2014).

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. R package version, 2(1), 74.

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. https://doi.org/10.1006/ccog.1997.0325

Brasil-Neto, J. P., Pascual-Leone, A., Valls-Solé, J., Cohen, L. G., & Hallett, M. (1992). Focal transcranial magnetic stimulation and response bias in a forced-choice task. *Journal of Neurology, Neurosurgery & Psychiatry, 55*(10), 964–966. https://doi.org/10.1136/jnnp.55.10.964

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(2), 121–138. https://doi.org/10.1017/S0140525X09000545

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351*(1346), 1413–1420.

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

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), 577–592. https://doi.org/10.1037/0096-3445.136.4.577

Duncker, K. (1945). On problem-solving. (L. S. Lees, Trans.). *Psychological Monographs, 58*(5), i–113. https://doi.org/10.1037/h0093599

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences, 9*(2), 127–152. https://doi.org/10.1007/BF00143739
Grant, E. R., & Spivey, M. J. (2003). Eye Movements and Problem Solving: Guiding Attention Guides Thought. *Psychological Science, 14*(5), 462–466. https://doi.org/10.1111/1467-9280.02454

Grimmer, H., Laukkonen, R., Tangen, J. M., & Hippel, B. von. (2021). Eliciting false insights with semantic priming. PsyArXiv. https://doi.org/10.31234/osf.io/kwhe9

Hattori, M., Sloman, S. A., & Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin & Review, 20*(4), 790–797. https://doi.org/10.3758/s13423-013-0389-0

Hedne, M. R., Norman, E., & Metcalfe, J. (2016). Intuitive feelings of warmth and confidence in insight and noninsight problem solving of magic tricks. *Frontiers in Psychology, 7*, 1314.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. *Science, 310*(5745), 116–119.
https://doi.org/10.1126/science.1111709

Klein, G., & Jarosz, A. (2011). A Naturalistic Study of Insight. *Journal of Cognitive Engineering and Decision Making, 5*(4), 335–351. https://doi.org/10.1177/1555343411427013

Koltko-Rivera, M. E. (2004). The Psychology of Worldviews. *Review of General Psychology, 8*(1), 3–58. https://doi.org/10.1037/1089-2680.8.1.3

Laukkonen, R. E., Kaveladze, B. T., Tangen, J. M., & Schooler, J. W. (2020). The dark side of Eureka: Artificially induced Aha moments make facts feel true. *Cognition, 196*, 104122.

Laukkonen, R. E., Ingledew, D. J., Grimmer, H. J., Schooler, J. W., & Tangen, J. M. (2021). Getting a grip on insight: real-time and embodied Aha experiences predict correct solutions. *Cognition and Emotion, 1*-18.

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

Laukkonen, R. E., & Tangen, J. M. (2018). How to detect insight moments in problem solving experiments. *Frontiers in psychology, 9*, 282.

Laukkonen, R., Webb, M. E., Salvi, C., Tangen, J. M., & Schooler, J. (2018). Eureka Heuristics: How feelings of insight signal the quality of a new idea [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ez3tn

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

Maier, N. R. (1930). Reasoning in humans. I. On direction. *Journal of Comparative Psychology, 10*(2), 115.

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12*(2), 181–194. https://doi.org/10.1037/h0071361
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution, 4(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Nasar, S. (2001). A Beautiful Mind: The Life of Mathematical Genius and Nobel Laureate John Nash. Simon and Schuster.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231.

Ohlsson, S. (1984). Restructuring revisited.: I. Summary and critique of the Gestalt theory of problem solving. Scandinavian Journal of Psychology, 25(1), 65–78. https://doi.org/10.1111/j.1467-9450.1984.tb01001.x

Reber, R., & Schwarz, N. (1999). Effects of Perceptual Fluency on Judgments of Truth. Consciousness and Cognition, 8(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Salvi, C., & Bowden, E. (2020). The relation between state and trait risk taking and problem-solving. Psychological Research, 84(5), 1235–1248. https://doi.org/10.1007/s00426-019-01152-y

Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., & Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. Psychonomic Bulletin & Review, 22(6), 1814–1819. https://doi.org/10.3758/s13423-015-0845-0

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. Thinking & Reasoning, 22(4), 443–460. https://doi.org/10.1080/13546783.2016.1141798

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight.

Schunn, C. D., & Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. Memory & Cognition, 24(3), 271–284. https://doi.org/10.3758/BF03213292

Schwarz, N. (2012). Feelings-as-Information Theory. In P. Van Lange, A. Kruglanski, & E. Higgins, Handbook of Theories of Social Psychology: Volume 1 (pp. 289–308). SAGE Publications Ltd. https://doi.org/10.4135/9781446249215.n15

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. Advances in Experimental Social Psychology, 39, 127–161.

Semmler, C., & Brewer, N. (2002). Effects of mood and emotion on juror processing and judgments. Behavioral Sciences & the Law, 20(4), 423–436.
Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135*(1), 94–120. https://doi.org/10.1037/a0014212

Sips, R. (2019). Psychosis as a Dialectic of Aha- and Anti-Aha-Experiences. *Schizophrenia Bulletin, 45*(5), 952–955. https://doi.org/10.1093/schbul/sby072

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

Topolinski, S., & Reber, R. (2010). Gaining Insight Into the “Aha” Experience. *Current Directions in Psychological Science, 19*(6), 402–405. https://doi.org/10.1177/0963721410388803

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

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

Wegner, D. M. (2004). Précis of The illusion of conscious will. *Behavioral and Brain Sciences, 27*(5), 649–659. https://doi.org/10.1017/S0140525X04000159

Wegner, D. M., & Wheatley, T. (1999). Apparent Mental Causation. *American Psychologist, 13*.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of experimental psychology. General, 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

**Figures**
Figure 1

Going from left to right, in each Anagram trial participants were presented with an incomplete claim, for example: “Free will is a powerful ____”. Below it was an anagram that completed the claim (e.g., “oinliusl”). When the answer to the anagram was submitted in the text entry box, or the visible 15-second timer ran out, participants were advanced to the next page. On that page, participants saw the completed claim: “Free will is a powerful illusion” and were asked to make a truth judgment about the claim. Participants then reported whether they experienced an aha moment while solving the anagram.
Figure 2

Combined data of the confirmation and the replication, N = 2,489. The plots (including means and standard deviations) illustrate truth ratings for the different conditions and key comparisons, with p-values and effect sizes. Each large black dot represents the mean and black lines represent +/- 1 standard deviation. Circles represent individual participants with a random horizontal jitter to aid visualization.
Figure 3

The plots in the top figure illustrate truth ratings for the four different between-subjects conditions, as well as the key analysis comparing anagrams without a delay (Anagram Normal condition) to anagrams with a delay (Anagram After Delay condition). In the bottom of the figure, the conditions are split between presence or absence of aha, illustrating an effect of aha on truth only when the anagram is presented at the same time as the worldview belief. Each large black dot represents the mean and black lines represent +/- 1 standard deviation. The circles represent individual participants with random horizontal jitter to aid visualization.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx