The Relationship between Creativity and Insight: A Case of the Wrong Answer?

Wendy Ross

Department of Psychology,

London Metropolitan University.

Address correspondence to Wendy Ross, Department of Psychology, London Metropolitan University, London, N7 8DB. w.ross@londonmet.ac.uk

Acknowledgements

These data were first presented at a symposium at the BPS Cognitive and Developmental Sections Conference in Bristol in 2023 and I thank the other presenters and the audience for insightful comments and feedback. I also thank Tom Ormerod for detailed comments on a previous version of this manuscript.

Abstract

Creative thinking is considered to be a core skill for the 21st century, specifically, the skills underlying creative problem-solving. Many studies in creative problem-solving rely on the problem-solvers using novel means to generate the correct answers. The current exploratory study draws on a secondary data to demonstrate that the wrong answers to problems can be just as creative as the correct answer although they are accompanied by lower feelings of insight. Categorising different types of wrong answers could be useful to provide a more granular assessment of the different cognitive processes underlying creative thinking.

The Relationship between Creativity and Insight: A Case of the Wrong Answer?

According to the World Economic Forum's Future of Jobs Survey 2023, creative thinking is the core skill growing most rapidly in importance. Underlying this survey is the view of creative thinking as being closely aligned with a *problem-solving* approach (WEF Skills Taxonomy); creativity is seen as an essential skill to solve the complex problems that society faces. From an academic perspective, therefore, identifying and nurturing the skills which underlie creative problem solving should be key to develop strategies to deal with an uncertain future.

Implicit theories around of creative problem-solving focus on the notion of a light bulb or a sudden illumination (Glăveanu, 2011) marking the sharp realisation of the correct answer. This form of creative problem solving is commonly referred to as insight problem solving, broadly defined as a sudden clarity about a situation with little or no conscious awareness of how that understanding was attained(Martínez-Ordaz, 2023). The lack of clear process for attaining this new understanding makes it of acute interest to those interested in novel thoughts and the site of different theoretical explanations and debate (Vallée-

Tourangeau, 2018). Most often, cognitive psychologists use simple or toy problems as a vehicle to explore this phenomenon (although see Hill & Kemp, 2018 for a more ecologically valid perspective). Such problems are constructed so that they clearly elicit the phenomenon of interest – that is a *feeling* of insight – caused by a restructuring of the problem space and a break with the old unproductive ideas of thinking (Webb et al., 2019; Weisberg, 2018). This feeling of insight is, therefore, identified by widely recognised phenomenological markers such as surprise, happiness and confidence that the answer that has been selected is the correct one (Danek et al., 2014; Jung-Beeman et al., 2004).

Importantly, the use of problem-solving stimuli with clear normatively correct answers has led to an association between solving an insight problem and being creative. This elision is not unjustified: as I have outlined above the idea of a light bulb moment marking a moment of creative thinking is ubiquitous across most Western cultures. It also reflects the dominant view that creativity operates best as blue sky thinking and that creative solutions require the problem-solver to remove the previous constraints (Montuori & Purser, 1995; Tromp, 2022). However, as creative problem-solving is being employed to solve more complex problems and as our understanding of the dominant view of creativity is becoming more nuanced, the link between the two bears some examination.

Previous research has shown that with more complex problems that are not deliberately structured to require a re-representation of the problem space, the link between a correct answer and the feeling of insight is less strong (Danek & Salvi, 2020). It is possible to experience and elicit false insight (the feeling of insight attached to an incorrect answer; Grimmer et al., 2022) and insight appears to be linked to confidence rather than objective correctness (Ross & Vallée-Tourangeau, 2022). Moreover, there is increasing evidence that insight-type problems can be solved through analytic and incremental means specifically drawing on previous knowledge (Ormerod et al., under review). If research in insight

problem solving is to support broader applications of creative problem-solving skills, it is important to understand how the hypotheses processes relate when the problems are scaled up in importance.

The Current Study

The current study is a secondary data analysis of data reported in Ross and Vallée-Tourangeau (2022). Participants were invited to solve a series of stumpers. A stumper is defined as "a riddle the solution to which is typically so elusive that it does not come to mind, at least initially - leaving the responder stumped" (Bar-Hillel, 2021, p. 1). For example, consider the following. "A big brown cow is lying down in the middle of a country road. The streetlights are not on, the moon is not out, and the skies are heavily clouded. A truck is driving towards the cow at full speed, its headlights off. Yet the driver sees the cow from afar easily, and avoids hitting it, without even having to brake hard. How is that possible?" (Bar Hillel et al., 2019, p. 112). The problem is difficult because the mention of streetlights, moon and headlights leads the solver to generate a mental representation of a nighttime scene. The answer involves dropping this and realising that the incident is taking place in the daytime.

Stumpers have a broader range of possible answers that could be suggested by participants - the appendix to Bar-Hillel et al. (2019) lists various inventive responses which were technically 'correct' with a level of logical contortion even if they were not the normative answer. In this way, stumpers offer the possibility which is not present in many insight tasks of collating and analysing plausible wrong answers to these forms of problems. This exploratory study assesses the relationship between creativity, insight and correctness for one stumper which runs thus:

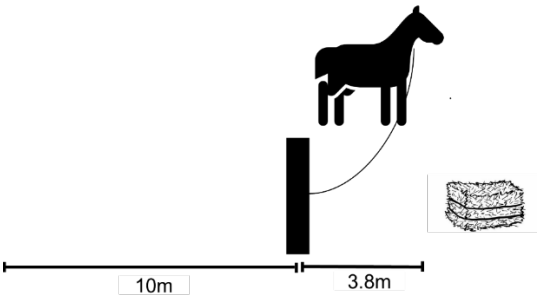
> A hungry horse is tied by its neck to a 10-metre-long chain. A bale of hay is 13.8 m away from it. Explain briefly how the horse reaches the hay with the chain intact.

The expected or normatively correct answer to the question is that the other end of the chain is not attached to anything. This answer relies for its effectiveness on the stumper originally directing the reader to generate a misleading representation of the problem space (the chain is attached on both ends) and this solution requires the participant to drop the initial assumptions.

However, this riddle also has two other answers which would also be technically correct without violating the problem statement or natural laws. First is the introduction of another agent (such as a farmer) to move the bale of hay closer to the horse. Second, the horse could be tethered to a post which is 10 metres away from where it is now but only 3.8 metres away from the bale of hay as in Figure One. Both these other correct answers do not require a restructuring of the problem statement but instead involve a different solution pathway. The first requires the participant to ignore the implicit rules of the experimental situation that a riddle requires a trick of some sort and the second requires both ignoring this rule and also having a level of spatial intelligence to allow a clear visualisation of the problem solution. Alongside this there are other answers which either violate the problem statement (for example that the horse bites through the chain when the problem states that the chain remains intact) or natural laws (for example the horse has a long neck) which can be more properly considered the wrong answer.

Figure One

The "Spatially Astute" Solution



There are more categories of non-expected correct answers than correct answers – the possibility spaces for failures is far greater than for success (Ormerod, 2023). The non-expected correct answers do not require the swift, all-or-nothing restructuring of the problem-space and should therefore be accompanied by lower feelings of insight. However, if insight is a marker of perceived correctness of the idea this may not be the case. Therefore, the first hypothesis is that there will be variation in the levels of the feeling of insight reported. Second, for the use of correct answers to reflect creative thinking, the non-expected correct answers should be rated lower in creativity.

Method

Participants

Data were retained from 148 participants in Experiment One and 147 participants in Experiment Two. 45% of participants in Experiment One were women, 51% were men and 4% did not identify as men or women. There was an average age of 26.96 years (*SD* = 6.05). In Experiment Two, 81% of participants were women, 17% were men and 2% did not identify as men or women. The average age for this experiment was 27.11 (*SD* = 9.29).

Stimuli

The data reported in this paper were drawn from the answers to one stumper which runs thus:

> A hungry horse is tied by its neck to a 10-metre-long chain. A bale of hay is 13.8 m away from it. Explain briefly how the horse reaches the hay with the chain intact.

Qualitative Coding

The set of answers was read entirely by WR who generated seven different categories of answers as laid out below. Two members of the research team who were blind to the hypotheses coded the answers into these categories. The raters agreed on 92% of codes and after discussion consensus was reached on all answers. The wrong answers and a representative correct answer (in total 171 responses) were then coded for creativity by 3 separate Mechanical Turk workers who were paid $0.03 per response.

They were instructed to code using the following wording which is adapted from standard instructions for the AUT task (see Silvia et al., 2008):

> These are answers to the question “A hungry horse is tied by its neck to a 10-metre-long chain. A bale of hay is 13.8 m away from it. Explain briefly how the horse reaches the hay with the chain intact.” Please rate each answer on how creative it is rather than whether it is correct or plausible. For example, the answer "If the horse is a sci-fi aficionado, it could dig a tunnel 3.8 meters long at an angle such that when it emerges, it is next to the hay. All while being tied by the chain, of course." would be related as highly creative (5). Creativity is defined in this task as ideas that strike people as clever, unusual, interesting, uncommon, humorous, innovative, or different. The ideas don't have to be practical or realistic; they can be silly or strange, even. There is no right or wrong answer, it is based on your opinion.

The interrater reliability was poor, $\kappa = .13$, which indicates the difficulty in fully rating for creativity for this task. As the nature of Mechanical Turk outsourcing precluded discussion an average creativity rating was generated across all three raters.

Results

Types of Wrong Answer

Table One

The Proportion of Answers Coded Into Each Category and The Relative Levels of Aha and Rated Creativity as a Function of Those Categories

	n	Proportion	Aha		Creativity	
			M	*SD*	*M*	*SD*
Correct	125	0.42	43.81	29.42	3.67	*NA*
Feet/Kicking	24	0.08	18.00	13.48	3.33	0.54
Impossible Body	35	0.12	21.24	19.49	3.31	0.72
Omission	18	0.06	-	-	-	-
Other	49	0.17	21.82	23.57	2.78	0.88
Other Agent	27	0.09	22.35	23.38	2.63	0.66
Spatially Astute	17	0.06	33.08	23.86	3.67	0.90

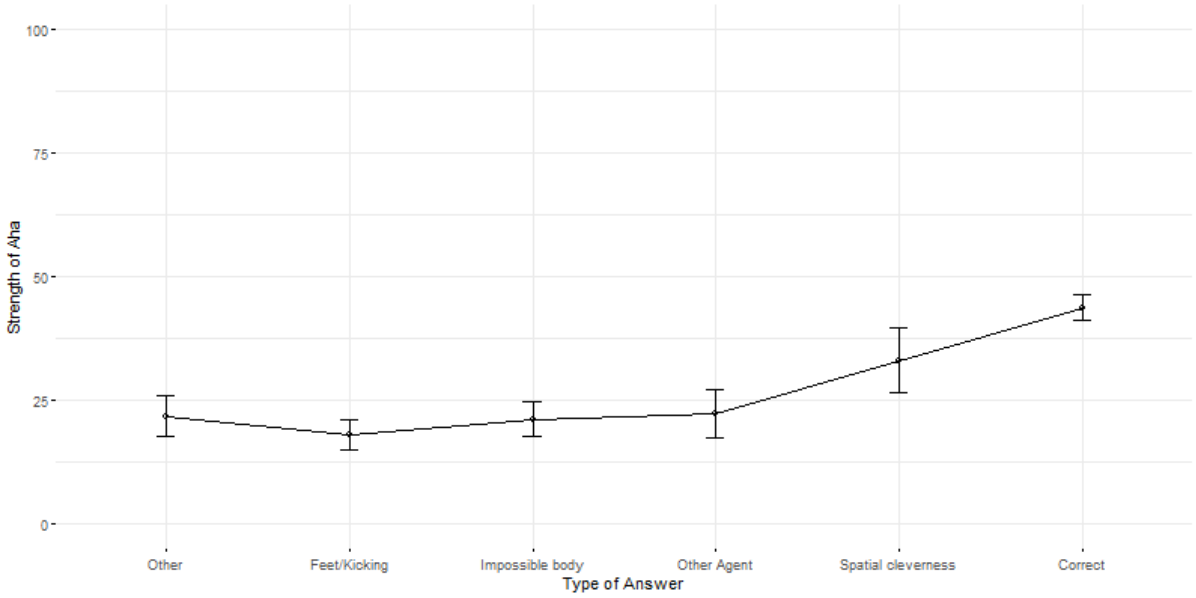
As can be seen in Table One, the majority of participants (42%) generated the correct answer or one which did not fall easily into one of the other categories (17%). 12% violated natural laws in some way such as giving the horse an impossibly constructed body (12%) or allowing it supernatural kicking power (8%). However, 9% of participants introduced another agent and 6% of participant generated the answer which relies on spatial awareness.

Levels of Aha

Participants were invited to rate the feeling of "Aha" on generating an answer and before they were told if they had the correct answer. Those who failed to generate an answer at all were not asked their levels of "Aha" and so are omitted from this analysis.

Figure Two

Strength of Aha as a Function of Type of Answer (Error Bars Indicate SEM)



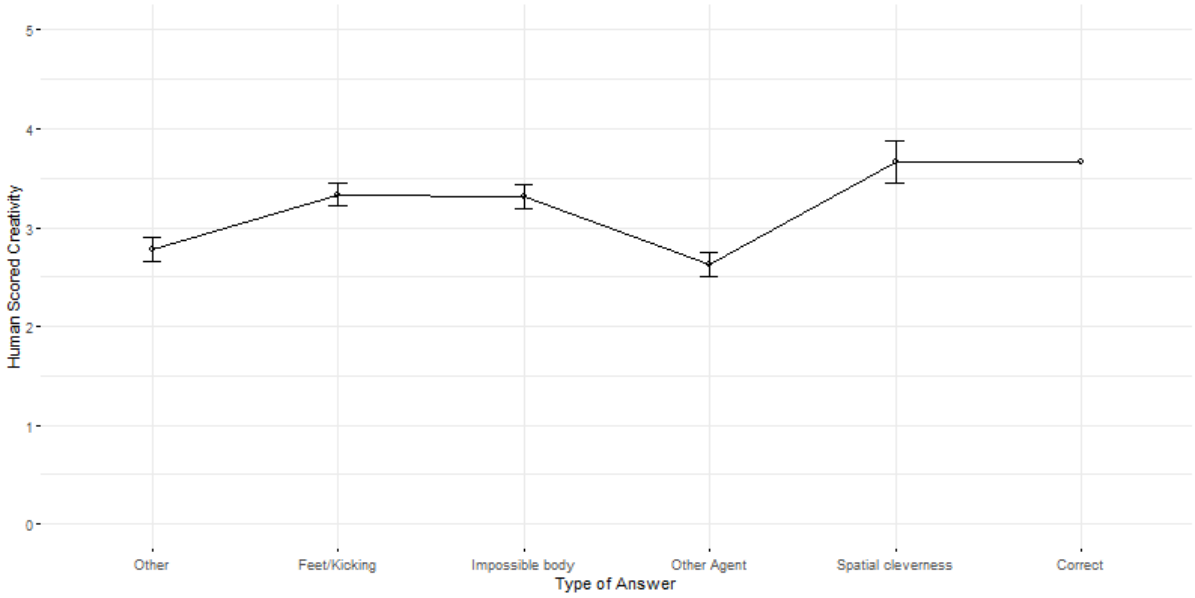
As can be seen in Figure Two and Table One, the normatively correct answer generated the strongest feeling of "Aha". We fitted a linear model (estimated using OLS) to predict Aha with answer type. The model explains a statistically significant and moderate proportion of variance, $R^2 = 0.16$, $F(5, 231) = 8.83$, $p < .001$, adj. $R^2 = 0.14$. All categories were significantly different to the correct answer in terms of the strength of "Aha" that they elicited (all p <.001) except the answer for the spatially astute which did elicit a weaker feeling of insight but not significantly so, $\beta = -10.74$, 95% CI [-25.53, 4.06], $t(231) = -1.43$, $p = .154$; Std. $\beta = -0.39$, 95% CI [-0.92, 0.15].

Levels of Creativity

As can be seen in Figure Three and Table One, the average creativity rating for the answer relying on spatial astuteness (M = 3.76, SD =0.9) was exactly the same as the score given to the expected answer.

Figure Three

Average Creativity as a Function of Type of Answer (Error Bars Indicate SEM)



A linear model was fitted to predict creativity with answer category. The model explains a statistically significant and moderate proportion of variance, R^2 = 0.19, $F(5, 147) = 6.68$, $p <$ .001, adj. R^2 = 0.16). There are no significant differences in creativity rating between the answer types (smallest p [Other Agent] = .184).

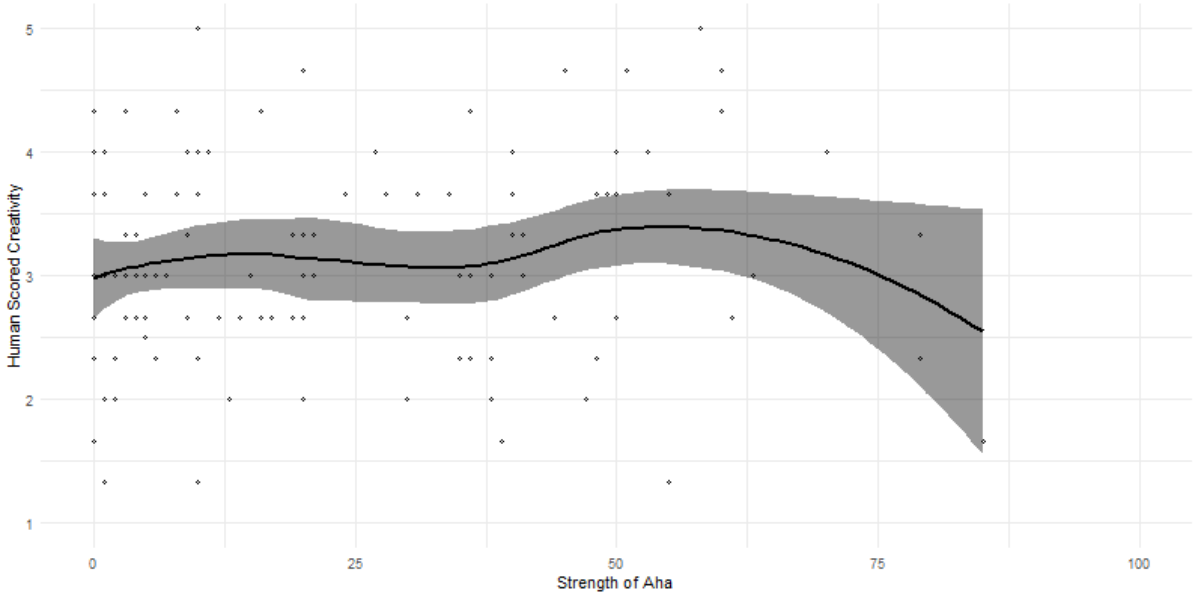
The Relationship between Insight and Creativity

Figure Four illustrates the relationship between the feeling of insight and the creativity of the answer. There is no significant relationship between the two; a linear model (estimated using

OLS) to predict creativity with strength of aha explains a statistically not significant and very weak proportion of variance, $R^2 = 0.01$, $F(1, 117) = 1.38$, $p = 0.243$, adj. $R2 < .001$). The effect of the strength of insight is statistically non-significant and positive ($\beta = 0.00415$, 95% CI [-0.00285, 0.01], $t(117) = 1.17$, $p = .243$; Std. $\beta = 0.11$, 95% CI [-0.07, 0.29])

Figure Four

The Relationship Between the Strength of Insight and Creativity (Shaded Area Indicates 95% Confidence Intervals)



Discussion

The feeling of insight was strongest for the normatively correct answer but although it was lower, it was not significantly lower for the answer relying on spatial intelligence. There were no significant differences in creativity scores across the different categories of answers whether they yielded a correct and plausible solution or not. There was no significant link between levels of insight and the rated creativity of the answer.

The current findings indicate that we should be careful about making a clear link between creativity and generating the expected answers to an insight problem. Or indeed, between the feeling of insight and the need for a restructuring. That the feeling of insight was lower for the non-expected correct answers but not significantly so for at least one of them suggests that there is not a direct link between the feeling of insight and mental restructuring. A normatively correct answer may elicit a stronger feeling of insight and adhere to the *rules* of the insight game, but it is not linked to the creativity or otherwise of the proposed solution. Moreover, answers which rely on a more systematic approach to the problem seemed to elicit a feeling of insight based not on their correctness (both the other agent and spatially clever answers were correct) but on another characteristic – perhaps their parsimony.

In addition, the creativity of a problem solution is also be constrained by the requirement to solve the problem. In the case outlined here, the answer that garnered the lowest creativity scores – employ another agent – is actually the answer which would be most useful were the participants to actually face the problem that was outlined (Ross & Vallée-Tourangeau, 2021). Although this requires an additional agent so appears at first glance to be less useful, the proposed answer, that the end of the rope is not tied would not be considered creative – indeed, it would probably be seen as being a facetious response as would the response that the horse is tethered to a post only 3.8 metres away from the hay. If these were the solutions there would be no actual problem to face. These two solutions, therefore, were only creative within the context of the riddle. Their creativity would shift if they were actually faced with a hungry horse.

In other words, current proposals that insight is linked to accuracy or creativity are undermined by the answers generated to this problem. While the assessment of creativity in this case suffered from the methodological problem of the poor level of interrater reliability, the general trends seemed stable. This poor level of inter-rater reliability could have been

because of the use of Mechanical Turk workers but it also suggests that it is not clear what is considered creative in this context despite the instructions. This may be because the raters were less experienced with rating answers to a problem rather than ideas which are less constrained in nature, such as to generate as many ideas as possible you can for a cardboard box in an Alternative Uses Task (Reiter-Palmon et al., 2019).

More broadly, researchers should be careful about the use of closed, simple problems which reliably elicit the feeling of insight to investigate complex phenomena such as creative thinking. The small amount of uncertainty about what would be correct in the answers given here should give researchers reasons to be reticent about moving too directly from the laboratory to more complex problem solutions.

References

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

Danek, A. H., & Salvi, C. (2020). Moment of truth: Why Aha! experiences are correct. *The Journal of Creative Behavior*, *54*(2), 484–486. https://doi.org/10.1002/jocb.380

Glăveanu, V. P. (2011). Is the lightbulb still on? Social representations of creativity in a western context. *The International Journal of Creativity and Problem Solving*, *21*(1), 53–72.

Grimmer, H., Laukkonen, R. E., Tangen, J., & von Hippel, W. (2022). Eliciting false insights with semantic priming. *Psychonomic Bulletin & Review*, *29*(3), 954–970. https://doi.org/10.3758/s13423-021-02049-x

Hill, G., & Kemp, S. M. (2018). Uh-Oh! What have we missed? A qualitative investigation into everyday insight experience. *The Journal of Creative Behavior*, *52*, 201–211. https://doi.org/10.1002/jocb.142

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural Activity When People Solve Verbal Problems with Insight. *PLoS Biology*, *2*(4), e97. https://doi.org/10/b3tpwf

Martínez-Ordaz, M. D. R. (2023). Scientific understanding through big data: From ignorance to insights to understanding. *Possibility Studies & Society*, *1*(3), 279–299. https://doi.org/10.1177/27538699231176523

Montuori, A., & Purser, R. E. (1995). Deconstructing the lone genius myth: Toward a contextual view of creativity. *Journal of Humanistic Psychology*, *35*, 69–112.

Ormerod, T. C. (2023). Possible, yes, but useful? Why the search for possibilities is limited but can be enhanced by expertise. *Possibility Studies & Society*, *1*(1).

Ormerod, T. C., MacGregor, J. N., Banks, A., & Rusconi, P. (under review). PRODIGI: A theory and computational implementation of insight problem-solving through progress maximization and idea discovery. *Psychological Review*.

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144–152. https://doi.org/10.1037/aca0000227

Ross, W., & Vallée-Tourangeau, F. (2021). Rewilding cognition: Complex dynamics in open experimental systems. *Journal of Trial and Error*, *2*.

Ross, W., & Vallée-Tourangeau, F. (2022). Insight with stumpers: Normative solution data for 25 stumpers and a fresh perspective on the accuracy effect. *Thinking Skills and Creativity*, 101114. https://doi.org/10.1016/j.tsc.2022.101114

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Tromp, C. (2022). Creativity From Constraint Exploration and Exploitation. *Psychological Reports*, *1*, 1–26. https://doi.org/10.1177/00332941221114421

Vallée-Tourangeau, F. (Ed.). (2018). *Insight: On the origins of new ideas.* Routledge.

Webb, M. E., Laukkonen, R. E., Cropper, S. J., & Little, D. R. (2019). Commentary: Moment of (Perceived) Truth: Exploring Accuracy of Aha! Experiences. *The Journal of Creative Behavior*, jocb.433. https://doi.org/10.1002/jocb.433

Weisberg, R. W. (2018). Insight, problem solving, and creativity: An integration of findings. In F. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas.* (pp. 191–215). Routledge.