

This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Published version: <https://doi.org/10.1016/j.cognition.2021.104946>

Aha! Under Pressure: The Aha! Experience is Not Constrained by Cognitive Load

Stuyck Hans^{(1,2)*}, Cleeremans Axel⁽²⁾ & Van den Bussche Eva⁽¹⁾

¹*KU Leuven, Faculty of Psychology and Educational Sciences, Brain and Cognition, Tiensestraat 102, 3000 Leuven, Belgium*

²*Université libre de Bruxelles, Faculty of Psychology and Education Sciences, Center for Research in Cognition and Neurosciences, Franklin Rooseveltlaan 50, 1050 Brussel, Belgium*

hans.stuyck@kuleuven.be

axcleer@ulb.ac.be

eva.vandenbussche@kuleuven.be

*Corresponding author: hans.stuyck@kuleuven.be

ABSTRACT

The Aha! moment— the sudden insight sometimes reached when solving a vexing problem— entails a different problem-solving experience than solution retrieval reached by an analytical, multistep strategy (i.e., non-insight). To date, the (un)conscious nature of insight remains debated. We addressed this by studying insight under cognitive load. If insight and non-insight problem solving rely on conscious, effortful processes, they should both be influenced by a concurrent cognitive load. However, if unconscious processes characterize insight, cognitive load might not affect it at all. Using a dual-task paradigm, young, healthy adults ($N = 106$) solved 70 word puzzles under different cognitive loads. We confirmed that insight solutions were more often correct and received higher solution confidence. Importantly, as cognitive load increased, non-insight solutions became less frequent and required more solution time, whereas insightful ones remained mostly unaffected. This implies that insight problem solving did not compete for limited cognitive resources.

Keywords: Insight, dual-task paradigm, creativity, consciousness, problem solving

1 INTRODUCTION

The "Aha! Experience" — that moment when the solution to a vexing problem suddenly pops into consciousness — has mesmerized scientists and laymen alike (Bowden & Grunewald, 2018; Chein & Weisberg, 2014). Scientifically, this singular subjective experience is referred to as *insight* (Bowden & Grunewald, 2018). While insight is not rare, most problem-solving involves a multistep analytical strategy (Simon & Newell, 1970; Weisberg, 2015) through which the problem solver searches long-term memory for potential algorithms, schemas, analogies or factual knowledge. This process is referred to as *analysis* or *non-insight* (e.g., Bowden, Jung-Beeman, Fleck, & Kounios, 2005; Fleck & Weisberg, 2013). It requires the problem solver to monitor his/her progress while maintaining the mental representation of the problem and avoiding being distracted by irrelevant information (see Shipstead, Harrison, & Engle, 2016; Wiley & Jarosz, 2012, for a review).

In contrast to insight, which feels effortless, non-insight solution-finding relies on continuous, effortful processing (e.g., Metcalfe & Wiebe, 1987). The difference between effortless and effortful processing also forms the core of dual-processing theories (Evans & Stanovich, 2013; Gilhooly, Ball, & Macchi, 2015; Sowden, Pringle, & Gabora, 2015). In such theories, Type 1 processes are assumed to be unconscious, effortless, and not limited by working memory (WM) limits. Type 2 processes, on the other hand, are taken to be conscious, effortful, and constrained by WM limits (Evans, 2019; Evans & Stanovich, 2013). For example, it has been shown that if participants have to make judgments while also performing an unrelated secondary task known to deplete cognitive resources, intuition-based judgments (Type 1) mainly stay unaffected, but deliberate-thought-based judgments are hampered (Type 2; De Neys, 2006; Howarth, Handley, & Walsh, 2016). This observation indicates that Type 1 processes are relatively independent of the cognitive resources needed to consciously manipulate information, whereas Type 2 processes depend more on cognitive resources (see Evans, 2019 for a review). Although it is intuitively appealing to consider insight resulting from Type 1 and non-insight from Type 2 processes, this is not a widely held claim (Benedek & Fink, 2019). For instance, Weisberg (2015) argues that insight, just like non-insight, mostly relies on the same effortful (Type 2) processes to retrieve a solution.

Can the conception of Type 1 and Type 2 processes proposed in dual-processing theories be translated to insight and non-insight problem solving? Dual-processing theories depart from a default-interventionist model (Evans, 2019; but also see De Neys, 2021; Mega & Volz, 2014). Type 1 processes entail default, automatic intuition-based responses resulting from automated processes (e.g., locating sounds), stereotypes, beliefs, or automated skills (e.g., reading; Howarth et al., 2016). Therefore, these intuition-based responses have been regarded as heuristically driven, helping humans to navigate life efficiently (e.g., mindlessly driving to work; Kahneman, 2011). However, sometimes cues in the environment can erroneously trigger an intuition-based response (e.g., stereotype-based judgments), defying the logic of the task at hand (De Neys & Pennycook, 2019). In such cases, it has been found that humans tend to make wrong intuition-based choices by failing to override them by deliberate thought (Type 2) to solve the task (Frey, Johnson, & De Neys, 2018). Humans often opt for the path of least resistance and follow their error-prone intuitions by default, even though these can be wrong (Evans, 2019).

At first sight, the conception that Type 1 processes lead to error-prone intuitions seems perpendicular to how insight solutions have been perceived. For instance, Salvi, Bricolo, Kounios, Bowden, and Jung-Beeman (2016) showed across different types of problems that insight solutions were more often correct than non-insightful ones. Numerous other studies have corroborated this higher solution-accuracy effect of insight (e.g., Danek, Fraps, von Müller, Grothe, & Öllinger, 2014; Hedne, Norman, & Metcalfe, 2016; Webb, Little, & Cropper, 2016). However, the role of intuitive, yet error-prone, Type 1 processes might be different for different phases of the insightful solution search (Zander, Öllinger, & Volz, 2016).

Consistent with representational change theory on insight (Ohlsson, 1992, 2011), heuristically driven intuitions (Type 1) might initially mislead the problem solver to an ill-defined problem representation, similar to what is proposed in dual-processing theories (Zander et al., 2016). This ill-defined problem representation leads the problem solver astray as content activated from semantic memory will inevitably be insufficient to solve the problem, leading to an impasse (Öllinger & von Müller, 2017). This impasse serves a pivotal function as it propagates negative feedback through the information processing system, hereby decreasing the activation level of the ill-defined problem representation and redistributing the unconscious spreading of activation to more remote, yet unnoticed concepts in semantic memory (Ohlsson, 2011; Öllinger & von Müller, 2017). When this spread of activation converges on the vital concept needed to rectify the initial erroneous problem representation, the solution pops into consciousness as the complete solution path is revealed all at once (i.e., restructuring; Bowden, 1997; Ohlsson, 2011; Öllinger & von Müller, 2017). At this final stage, it is assumed that the problem solver is consciously puzzled but unaware of the unfolding spreading of activation through semantic memory (Ohlsson, 2011). Recently, it has been argued that the problem solver uses the Aha! phenomenology (i.e., positive affect, surprise, obviousness, high confidence) as a heuristic that signals the quality of a solution emerging from the unconscious (Laukkonen, Webb, Salvi, Tangen, & Schooler, 2018). For example, Laukkonen Kaveladze, Tangen, and Schooler (2020) found that false statements judged on their veracity are considered more true when linked with an Aha! experience than when judged without it (see also Laukkonen, Ingledew, Grimmer, Schooler, & Tangen, 2021). In that sense, Type 1 processes play a double part in insight problem-solving: although they may mislead the problem solver at first, in the end, they signal the quality of the solution surfacing from the unconscious via an Aha-heuristic (Laukkonen et al., 2018; Zander et al., 2016). This idea corresponds with recent research showing that Type 1 processes can culminate into accurate logic-driven intuitions (Bago & De Neys, 2019, see also Bowers, Regehr, Balthazard, & Parker, 1990) and experts' intuitions that can be dead right (see Salas, Rosen, & DiazGranados, 2010, for a review). These observations indicate that the error-proneness of Type 1 intuitions do not need to be irreconcilable with insight's high accuracy, and that Type 1 and Type 2 processes are useful concepts when thinking about insight and non-insight problem solving, respectively.

However, not all insight theorists agree with these assumptions (Benedek & Fink, 2019, Weisberg, 2015, 2018, see also Chater, 2018). Weisberg (2015) assumes that effortful Type 2 processes mainly dominate both insight and non-insight solution finding. Insight is assumed to arise via the problem solver's ability to capitalize on repeated failures to find the solution non-insightfully. The repeated failures bring forth new information to work with, guiding the solution search to new, promising directions that unveil ill-held assumptions about the problem. Once these ill-held assumptions are rectified (i.e., restructuring), the complete solution path is revealed all at once, and the solution is found with insight. Although Weisberg (2015) argued that insight might also be achieved via unconscious processes, he perceived this as a rare phenomenon (Weisberg, 2018).

As the above-theoretical views indicate, the debate concerning insight's (un)conscious nature is not settled yet. One way of debunking whether insight relies on cognitive resources is by assessing how WM relates to insight and non-insight problem solving (e.g., DeCaro, Van Stockum, & Wieth, 2016). WM is considered a central processing hub where information is shortly stored and updated to cope with ongoing task demands (Baddeley, 1986; Shipstead et al., 2016); executive functions associated with problem solving (e.g., Cooper & Marsh, 2015; De Neys, 2006). The capacity of WM is limited (Cowan, 2010). Some have defined this limited capacity in terms of the number of informational chunks that can be hold in the scope of attention (for a review, see Cowan, 2010; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), whereas others have conceived it as being more closely related to the limits of attentional control processes (e.g., filtering efficiency; see Oberauer, 2019, for an overview).

Thus, WM capacity imposes limits to our ability to maintain and update the mental representations of a problem (Shipstead et al., 2016; Wiley & Jarosz, 2012). Such limits should then clearly influence one's

ability to solve complex, multistep problems through non-insight. But what about insight? If insight is indeed associated with unconscious, Type 1 processes, it should not be bound by the limits of WM capacity (Fleck, 2008; Gilhooly & Fioratou, 2009). However, if it depends on conscious, effortful Type 2 processes, it should also be constrained by WM capacity limitations (Chein & Weisberg, 2014; Chuderski & Jastrzebski, 2018). One way of testing these assumptions consists of using a dual-task paradigm where participants are asked to perform a primary problem-solving task while concurrently executing a secondary WM task (e.g., retaining a series of digits; De Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012). Loading WM in such a way should therefore decrease its capacity and hence its availability to executive functions. Thus, if the primary task also involves WM, then concurrently executing the WM task should hamper performance on the primary task (e.g., Camarda et al., 2018). In contrast, if the primary task does not tax WM, then concurrently executing a WM task should not impact performance (e.g., Abadie, Waroquier, & Terrier, 2013).

A few studies have already explored the role of WM in problem solving. Lavric, Forstmeier, and Rippon (2000) presented participants with classical insight and non-insight problems. Single- and dual-task conditions were created by adding a concurrent tone-counting task for some participants but not for others. The insight problems were correctly solved by a comparable number of participants in both conditions. However, the non-insight problems were correctly solved by a larger number of participants in the single-task conditions, as compared to the dual-task condition. This finding is consistent with the assumption that insight, unlike non-insight, depends on Type 1 processes. While other studies have found similar results (e.g., Korovkin, Vladimirov, Chistopolskaya, & Savinova, 2018; Korovkin & Savinova, 2014), De Dreu et al. (2012), however, found that the number of correctly solved insight problems decreased as the concurrent WM load increased. The authors concluded that insight problem solving depends on WM and hence competes for limited WM capacity. This study and other studies (see also Lin & Lien, 2013; Wieth & Burns, 2014 for similar results) suggests that insight, like non-insight, relies on Type 2 processes.

The inconclusive nature of these findings might originate from their dependence on classical insight and non-insight problems and on the assumption that these problems reliably index insight and non-insight, respectively. However, these two types of problems are very different, making it difficult to draw strong conclusions based on their comparison (Bowden & Jung-Beeman, 2007; Webb et al., 2016). Moreover, it has been shown that insight problems are sometimes solved with non-insight and non-insight problems with insight (Danek, Wiley, & Öllinger, 2016; Webb et al., 2016). Therefore, studies using classical insight problems without a non-insight comparison group might have yielded confounding results (e.g., De Dreu et al., 2012). These findings indicate that insight and non-insight problem solving are difficult to pin down in a specific problem type. Therefore, some have suggested that the problem solver is the most designated person to decide how a solution was found because what sets insight apart from non-insight resides in the problem solver's phenomenology rather than the nature of the problem type (see Bowden & Grunewald, 2018).

To address these issues, researchers have developed problem types that have an almost equal likelihood of being solved with insight and non-insight, keeping the type of problems constant across both solution types. For example, in the compound remote associates test (CRA; Bowden & Jung-Beeman, 2003), participants receive three cue words (break/bean/cake) on each trial and are requested to search for a fourth compound solution word (*coffee* break/*coffee* bean/*coffee* cake). After each solved CRA trial, participants indicate whether they solved the problem with insight or non-insight based on their subjective, solution-finding experience (e.g., Salvi, Simoncini, Grafman, & Jung-Beeman, 2020).

However, this procedure to classify insight and non-insight solutions based on the participant's subjective report has not been without critique (Danek & Salvi, 2020; Laukkonen et al., 2021). The finding that insight solutions are more often correct, receive higher solution confidence, and are solved faster than

non-insightful ones (e.g., Hedne et al., 2016; Cranford & Moss, 2012) raises the question if those behavioral or phenomenological indices do not bias the retrospective insight classification. However, strong correspondence has been observed between participants' self-reported insights and physiological indices of insight (i.e., participants' squeeze strength on a dynamometer upon solving a problem; Laukkonen et al., 2021). Moreover, subjective self-reports of insight have been associated with distinct brain-pattern activation (i.e., activation burst across right temporal lobe; see Kounios & Jung-Beeman, 2014), physiological responses (i.e., increased skin conductance, heart rate, and pupil dilation; Salvi et al., 2020; Shen et al., 2017), and phenomenological qualia (feelings of happiness and relief; Stuyck, Aben, Cleeremans, & Van den Bussche, 2021). This body of research provides additional support for the validity of self-reports to study insight and non-insight problem solving.

In the current study, we, therefore, used a problem type (the CRA) that can be solved with insight and non-insight, which was determined based on participants' self-reports. We manipulated WM load by creating no-load, low-load, and high-load conditions. As non-insight problem solving is expected to be an effortful, conscious, Type 2 process, we predicted a detrimental influence of WM load on the performance of the problems solved with non-insight. If insight problem solving relies on an automatic, unconscious, Type 1 process, it should not be influenced by WM load. However, if insight relies on WM in the same way as non-insight, it should be impacted similarly by the WM load.

2 METHOD

2.1 Participants

A convenience sample of 106 psychology undergraduates of the KU Leuven participated in this study. They received course credit for their participation. The data of one participant were excluded due to issues with data acquisition. The final sample consisted of 105 participants, of which 91 were female. The mean age was 18 years ($SD = 0.72$, range 17-23). All participants had normal or corrected-to-normal vision. Ninety-eight participants had Dutch as their mother tongue, and seven were bilingual, with Dutch as their second language. We randomly assigned 30 participants to the no-load condition, 35 to the low-load condition, and 40 to the high-load condition. Before the start of the experiment, all participants provided written informed consent. The Social and Societal Ethics Committee of the KU Leuven approved the study (i.e., approval code G- 2018 10 1368).

Given the lack of previous studies using a dual-task paradigm with the CRA to study insight and using generalized linear mixed models analyses, we used the average effect size found in previous dual-task studies on insight to make a well-informed estimate of our required sample size. In previous dual-task studies, an average Cohen's d effect size has been reported of .68 for solution time on the secondary task and .62 for the number of correctly solved problems as a function of solution type and working memory load (Korovkin et al., 2018; Lavric et al., 2000). Regarding solution accuracy and solution confidence, we could not extract Cohen's d effect sizes due to a lack of previous data. Therefore, we a priori assumed a Cohen's d effect size of .4, generally considered a valid effect size in the research field of cognitive psychology (Brysbaert & Stevens, 2018). Given those medium-to-large effect sizes, we considered the current sample size as sufficient to detect effects with comparable size with a statistical power of 80%.

2.2 Material

This study used the Dutch version of the compound remote associates test (see <https://osf.io/snb3k/> for the Dutch CRA and selection procedure; Stuyck et al., 2021). The CRA consists of word puzzles that can be solved both with insight and non-insight. On each trial, participants are presented with three cue words (e.g.,

artist, hatch, route) for which they have to find a fourth compound solution word that they can attach to each of these three cue words (e.g., *escape* artist, *escape* hatch, *escape* route). Participants received 75 CRA trials, of which five were practice trials. The three cue words of the CRA and a question mark were always shown in vertical alignment at the center of a black screen (i.e., the upper word was placed at 25% of the screen's y-axis, each other word and finally the question mark were spaced with an additional 15% on that same axis). They were presented in white, in "Courier New" font with point size 18 and in bold typeface.

In the low-load and high-load conditions, participants additionally had to perform a WM task while solving the word puzzles. Similar to De Dreu et al. (2012), we asked participants to retain two digits in the low-load (e.g., 24) and four digits in the high-load (e.g., 9861) condition. Each two- or four-digit combination was unique and selected randomly. These digit combinations never contained zero or a duplication of a digit (e.g., 33 or 1244). Participants had to recall the digits in any order. In the high-load condition, performance on the WM task was deemed accurate if three-out-of-four digits were recalled correctly. The digits were always presented sequentially, in white at the center of a black screen, in "Courier New" font with point size 25 and in bold typeface.

2.3 Equipment

We tested participants in groups of maximum 12. They were seated at an individual computer in a quiet, daylight illuminated computer room. Sufficient space was created between participants so that they could not see each other's screen. The experiment was programmed with E-prime 2.0 (Psychology Software Tools; Schneider, Eschman, & Zuccolotto, 2002). For more technical information about the equipment, see <https://osf.io/sc5n7/>; www.pstnet.com.

2.4 Procedure

Figure 1 depicts an example of a CRA trial in the low-load condition. Participants initiated each trial by pressing the spacebar. In the *no-load condition*, a fixation cross appeared at the center of the screen for 2s. Subsequently, the CRA word puzzle was presented for a maximum of 25s. In the right upper corner of the screen, a countdown timer was presented. If the participants found the answer, they were instructed to press the spacebar as quickly as possible. They then received a screen prompting them to type in their solution. There was a time limit of 10s to type in their answer. After providing their answer, participants had to indicate their solution confidence (i.e., "please indicate on the scale below how confident you are in your solution?") by selecting a position on a horizontal visual analog scale with the cursor of their mouse within a time limit of 30s. This scale gradually changed from red (i.e., low confidence = 0) to green (i.e., high confidence = 100). The cursor of the mouse always started at the center of the scale. Next, participants had to indicate whether they had found the solution with insight, with non-insight or with another strategy by typing the number one, two, or three, respectively, without a time limit. During the instructions, participants had received the definition of each solution type based on previous studies on insight to ensure that participants clearly understood what each solution type entailed (see Appendix A; Danek et al., 2014; Hedne et al., 2016; Jung-Beeman et al., 2004). Moreover, instructions regarding the solution confidence rating were decoupled from the instructions of solution type (see Hedne et al., 2016, for a similar procedure) to reduce the likelihood that participants used solution confidence as a criterion to make the insight/non-insight classification. Participants of the no-load condition automatically continued to the next trial after indicating the solution type. If participants could not solve the CRA word puzzle within the allotted time, they automatically proceeded to the next trial. We presented the participants with five practice trials and 70 experimental trials, both in random order. After solving half of the experimental trials, a break was provided.

In the *dual-task conditions*, the procedure was identical, except for the following additions. Before each trial, participants were reminded that they first had to perform the WM task. This WM task's digits were then

presented sequentially at a pace of 1.5s per digit (i.e., low-load = two digits; high-load = four digits). After solving a word puzzle, participants were prompted to type in the digits that had to be remembered on that trial. There was no time limit to perform the WM task.

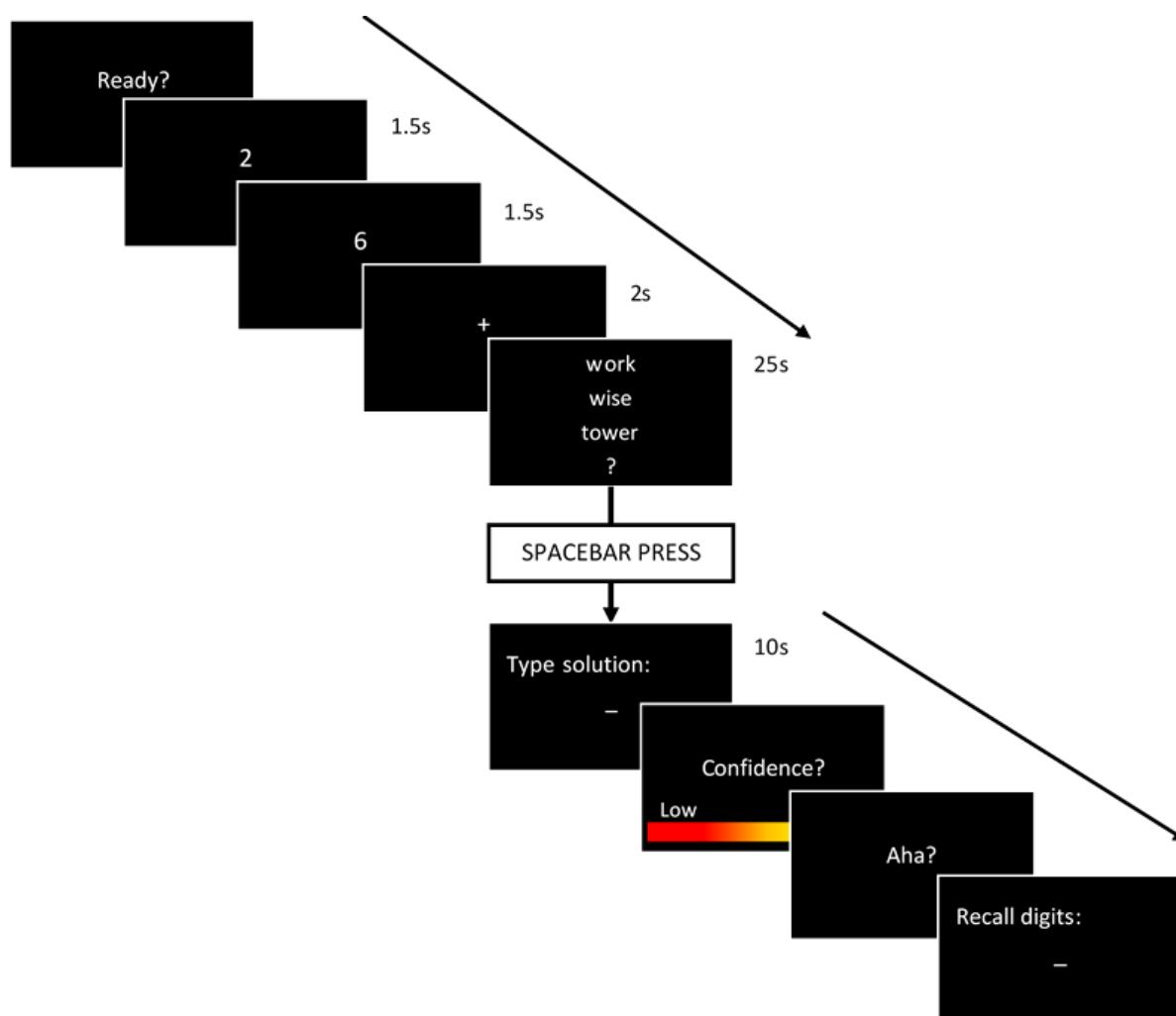


Figure 1. Example of a CRA trial for the low-load condition. In the high-load condition, participants received four digits instead of two. In the no-load condition, there was no WM task. The solution to this word puzzle is "clock".

2.5 Statistical analysis

For the statistical analysis, we only used the word puzzles that were solved with insight or non-insight. The third option, "another strategy", was excluded because it did not contain any information concerning the solution types. This omission led to the exclusion of 276 word puzzles out of the 4780 solved word puzzles. Furthermore, we only included solved CRA trials where the performance on the WM task was correct. This to ensure that participants were jointly performing both tasks adequately. Consequently, 315 solved CRA trials in the low-load and 431 in the high-load condition were excluded. The hypotheses and statistical analyses of this experiment were preregistered on the OSF platform (<https://osf.io/2v9yz>). Deviations from this preregistration can be found in Table 1 of Appendix B. All data and program code (R code and E-prime) were placed on the OSF platform (<https://osf.io/sc5n7/>).

To analyze the data, we used three (generalized) linear mixed models (i.e., [G]LMM) and one generalized linear model (GLM) with solution type (within-subject variable with two levels: insight and non-insight), WM load (between-subject variable with three levels: no-load, low-load and high-load) and their interaction term as fixed effects. A first LMM contained the log-transformed solution time (to accommodate the non-normality) of the correctly solved word puzzles as a continuous outcome variable (Gaussian error distribution; Baayen & Milin, 2010). A second GLMM contained solution accuracy (i.e., 0 = incorrect and

1 = correct) as a binary outcome variable (Binomial error distribution; Sommet & Morselli, 2017). A third GLMM contained solution confidence of the correctly solved word puzzles as a bounded outcome variable (i.e., transformed range .0050 to .9950; see Smithson & Verkuilen, 2006; Verkuilen & Smithson, 2012; Xu, Samtani, Yuan, & Nandy, 2014 for an in-depth explanation). To analyze the number of correctly solved word puzzles, the data were aggregated per participant into one observation for insight and one for non-insight per participant. This aggregated data had no trial-by-trial variation, contrary to the previous outcome variables. Hence, a GLMM with a complex random structure was unsupported (see Bates, Kliegl, Vasishth, & Baayen, 2018; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) and a generalized linear model (GLM) was used containing the number of correctly solved word puzzles as a count outcome variable (Negative Binomial error distribution; Gardner, Mulvey, & Shaw, 1995). To account for the non-independence of the data (i.e., two observations clustered within each participant), robust standard errors were used (for an in-depth explanation, see Zeileis, Köll, & Graham, 2020). The statistical significance of the fixed effects for the LMM was obtained by using the Satterthwaite approximation method. The Wald test assessed the statistical significance of the fixed effects of the GLMMs and the GLM.

The advantage of using (G)LMM is that participants and word puzzles can be treated as crossed random effects, thereby taking by-participant and by-word puzzle variations into account (Baayen, Davidson, & Bates, 2008). We included random intercepts for participants and word puzzles in all the (G)LMMs. Furthermore, using (G)LMMs, the data analysis can be performed at the observational level without the need to aggregate outcome variables by-participants or by-word puzzles (e.g., Judd, Westfall, & Kenny, 2012). All estimated models (i.e., the(G)LMMs and GLM) are provided in Appendix C.

The predictor variables solution type and WM load consist of discrete levels, namely two and three levels. Therefore we applied sum coding to set the contrasts for each estimated model (i.e., orthogonal contrasts; see Schad, Vasishth, Hohenstein, & Kliegl, 2020 for an in-depth explanation). Using sum coding to set the contrasts, it is possible to assess each estimated model's main effects and interaction effects in the same vein as is done in an ANOVA (see Levy, 2018, for an in-depth explanation). We examined the main effects and the interaction effect for each estimated model by comparing the full model, including all fixed effects, to three reduced models. Each reduced model excluded either one of the main effects (e.g., solution type or WM load) or the interaction effect. Subsequently, a likelihood ratio test for the (G)LMM (χ^2) and a Wald test for the GLM (χ^2) were used to measure if the full model explains more residual variance than the reduced models. Hence, finding a statistically significant effect implied a main effect or an interaction effect (Levy, 2018; Zeileis & Hothorn, 2002).

To explain the significant main and interaction effects of the estimated models, we used post-hoc tests. The post-hoc tests derive the estimated marginal means from each estimated model. In the case of a main effect, a pairwise contrast is made for either solution type (insight vs. non-insight) or WM load (all possible pairwise combinations) based on the estimated marginal means. If an estimated model contained an interaction effect, pairwise contrasts of the estimated marginal means for the three levels of WM load were made conditional on the solution types (i.e., no-load vs. low-load; no-load vs. high-load; low-load vs. high-load within insight and non-insight). The Tukey method was used to adjust for multiple comparisons. Cohen's *d* effect sizes are reported for significant pairwise contrasts. We interpreted the cohen's *d* effect size as follows: 0-0.09 = negligible, 0.10-0.30 = small, 0.31-0.39 = small-to-medium, 0.40-0.60 = medium, 0.61-0.69 = medium-to-large, and above 0.70 = large (see Pliatsikas et al., 2019 for a similar procedure).

We used the open-source *R* language and environment to perform statistical analysis (R Core Team, 2020). The (G)LMMs on solution time and solution accuracy were built with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). The GLMM on solution confidence was built with the *glmmTMB* package (Brooks et al., 2017). The GLM on the number of correctly solved word puzzles was built with the

MASS package (Venables & Ripley, 2002), and its robust standard errors were obtained with the *Sandwich* package (Zeileis et al., 2020). Post-hoc tests were performed with the *emmeans* package (Lenth, 2020).

2.6 Secondary analysis

A concern related to self-reports for the insight/non-insight classification is that participants might report having solved a word puzzle with insight because they found the solution quickly. It has been argued that such solutions result from immediate recognition of the solution rather than a problem-solving process (Cranford & Moss, 2012). Therefore, we performed a secondary analysis. In this secondary analysis, we reran the estimated models with the responses provided within the two first seconds excluded (see Salvi et al., 2020, for a similar procedure). This exclusion criterium led to the additional omission of 56 observations, of which 50 were correctly solved word puzzles. Under the assumption that participant's self-reported insight solutions are independent of these immediately recognized solutions, we expected to find a similar pattern of results for all estimated models when responses provided within the first two seconds are excluded. These results are presented in Appendix D.

3 RESULTS

After excluding word puzzles solved with "another strategy" ($N = 276$), with an incorrect solution of the WM task (low-load = 315 and high-load 431), and with an incorrect CRA solution ($N = 664$), the final sample of correctly solved word puzzles was 3094. Based on this final sample, the average number of correctly solved word puzzles with insight was 18 ($SD = 9$, range 1-40) and 12 with non-insight ($SD = 9$, range 1-39)¹. The descriptive statistics of the outcome variables and the total number of solved word puzzles are depicted in Table 1.

¹ The participant sample was not balanced for biological sex with 91 females and 14 males. To exclude any influence of biological sex on our results, we reconducted all statistical analysis with biological sex as covariate included in each estimated model. The pattern of results remained similar as well as the direction of the effects. For solution time, only a main effect of solution type ($p < .001$) was found and a trend towards an interaction effect ($p = .06$). For solution accuracy and solution confidence there only was a main effect of solution type ($p < .001$ and $p < .001$, respectively). For the number of correctly solved word puzzles, there was a main effect of solution type ($p < .001$), WM load ($p = .001$) and a significant interaction effect ($p = .006$). Therefore, we contend that the effect of biological sex is likely to be minimal (see also Wieth & Burns, 2006).

Table 1. The descriptive statistics of the outcome variables and the number of solved word puzzles

		No-Load		Low-Load		High-Load	
		M(SD)	range	M(SD)	range	M(SD)	range
Insight	solution time	3.80(0.26)	3.13-4.40	3.78(0.28)	3.11-4.39	3.77(0.25)	3.13-4.40
	confidence	.81(.19)	.16-.995	.80(.19)	.005-.995	.76(.19)	.005-.995
	#solved	18(8)	3-34	17(9)	2-39	19(9)	1-40
	accuracy		88%		91%		90%
	#correct		528		600		772
	#(in)correct		600		662		862
Non-Insight	solution time	3.84(0.27)	3.21-4.40	3.82(0.27)	3.14-4.39	3.84(0.26)	3.26-4.39
	confidence	.65(0.26)	.005-.995	.69(.27)	.005-.995	.63(.25)	.005-.995
	#solved	16(9)	2-39	12(10)	1-37	8(7)	1-29
	accuracy		71%		79%		67%
	#correct		474		421		299
	#(in)correct		664		534		436

Note. Solution time is log-transformed with base 10; confidence = solution confidence; solution confidence is rescaled to range from 0.005 = low solution confidence to 0.995 = high solution confidence; #solved = the average number of correctly solved word puzzles; accuracy = the mean percentage of correctly solved word puzzles; #correct = the total number of correctly solved word puzzles across all participants; #(in)correct = the total number of solved word puzzles regardless of correctness across all participants.

3.1 Solution time

An LMM was constructed with the (log-transformed) solution time of the correctly solved word puzzles as the outcome variable and solution type (two levels: insight and non-insight), WM load (three levels: no-load, low-load, and high-load) and their interaction term as fixed effects (see Table 1 in Appendix C for an overview). There was a significant main effect of solution type, $\chi^2(1) = 15.25, p < .001$. Word puzzles solved with insight ($M = 3.82$) were solved significantly faster than those solved with non-insight ($M = 3.86$), $t(2935) = -3.91, p < .001$, Cohen's $d = 0.17$ (95% CI [0.08, 0.25]). There was also a trend towards an interaction effect, $\chi^2(2) = 5.82, p = .055$. Post-hoc tests showed no significant differences between the three levels of WM load for insight (no-load vs. low-load, $p = .889$; no-load vs. high-load, $p = .523$ and low-load vs. high load, $p = .802$). A similar pattern of results was found for non-insight (no-load vs. low-load, $p = .791$; no-load vs. high-load, $p = .677$ and low-load vs. high load, $p = .303$). To clarify the interaction effect, pairwise comparisons were performed for the solution types conditional on the three levels of WM load. These post-hoc tests showed that only in the high-load condition, word puzzles solved with insight were solved significantly faster than word puzzles solved with non-insight ($M = 3.80$ versus $M = 3.88$), $t(2897) = -4.10, p < .001$, Cohen's $d = 0.32$ (95% CI [0.16, 0.47]). This was not the case in the no- and low-load conditions ($p = .131$ and $p = .261$).

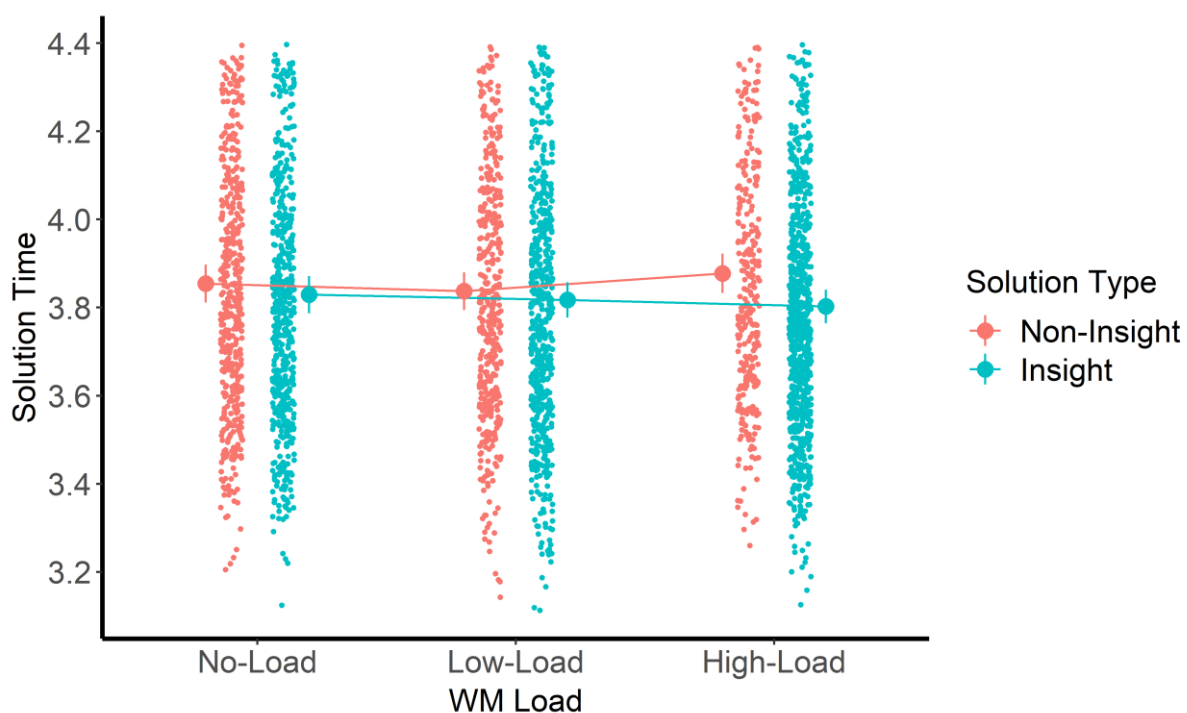


Figure 2. The interaction between solution type and WM load for solution time. Bars represent the 95% confidence intervals. Solution time was log-transformed with base 10.

3.2 Solution accuracy

Likewise, a GLMM was constructed with solution accuracy as a binary outcome variable (see Table 2 in Appendix C for an overview). There was a significant main effect of solution type, $\chi^2(1) = 197.13, p < .001$. Word puzzles solved with insight had a significantly higher probability of being correct than word puzzles solved with non-insight ($M = 94\%$ versus $M = 77\%$), $Z = 13.47, p < .001$, Cohen's $d = -1.57$ (95%

CI [-1.8, -1.34]). The main effect of WM load and the interaction effect were not significant, $p = .146$ and $p = .337$, respectively.

3.3 Solution confidence

Similarly, a GLMM was constructed with solution confidence of the correctly solved word puzzles as a bounded outcome variable (range .005-.995; see Table 3 in Appendix C for an overview). There was a significant main effect of solution type, $\chi^2(1) = 243.82$, $p < .001$. The word puzzles solved with insight received significantly higher solution confidence than word puzzles solved with non-insight ($M = .79$ versus $M = .66$), $t(3085) = 15.96$, $p < .001$, Cohen's $d = -0.15$ (95% CI [-0.17, -0.13]). The main effect of WM load and the interaction effect were not significant, $p = .085$ and $p = .605$, respectively.

3.4 The number of correctly solved word puzzles

Lastly, a GLM was built with the number of correctly solved word puzzles as a count outcome variable (see Table 4 in Appendix C for an overview)². There was a significant main effect of solution type, $\chi^2(3) = 27.29$, $p < .001$. Correctly solved word puzzles were more frequent for insight than non-insight ($M = 18.0$ versus $M = 11.4$), $Z = 5.04$, $p < .001$, Cohen's $d = -0.43$ (95% CI [-0.60, -0.26]). Also, a main effect of WM load was observed, $\chi^2(4) = 18.10$, $p = .001$. Post-hoc tests demonstrated that correctly solved word puzzles were more frequent in the no-load than in the high-load condition ($M = 16.7$ versus $M = 12.3$), $Z = -2.73$, $p = .018$, Cohen's $d = 0.29$ (95% CI [0.08, 0.50]). The comparison between the no- and low-load condition and between the low- and high-load condition were not significant ($p = .378$ and $p = .327$, respectively). More importantly, there was a significant interaction effect, $\chi^2(2) = 10.09$, $p = .007$. The post-hoc tests illustrated that, for insight, there were no significant differences between the three levels of WM load (no-load vs. low-load, $p = .985$; no-load vs. high-load, $p = .817$ and low-load vs. high load, $p = .696$). However, for non-insight, the number of correctly solved word puzzles were more frequent in the no- than high-load condition ($M = 15.8$ versus $M = 7.9$), $Z = -4.33$, $p < .001$, Cohen's $d = 0.66$ (95% CI [0.36, 0.97]), and more frequent in the low- than high-load condition ($M = 12.0$ versus $M = 7.9$), $Z = -2.71$, $p = .019$, Cohen's $d = 0.40$ (95% CI [0.11, 0.70]). The comparison between the no- and low-load condition was not significant ($p = .205$). Thus, as WM load increased the number of correctly solved word puzzles decreased for non-insight but remained unaffected for insight. This interaction is illustrated in Figure 3³.

² As the total number of correctly solved word puzzles varied among participants, we took this into account by including an offset in the GLM of the number of correctly solved word puzzles. Including this offset transforms the outcome variable to a rate (i.e., $\log(\text{number solved correctly}/\text{total number solved correctly})$) to estimate the model's parameters (see Hutchinson & Holtman, 2005). The results of this adjusted GLM remained similar to the initial analysis. Namely, a main effect of solution type ($p < .001$), a main effect of WM load ($p < .001$), and an interaction effect between solution type and WM load ($p = .008$). The direction of the effects remained similar to the initial analysis.

³ The above-depicted results are based on the observations for which participants had three-out-of-four correct recalled digits in the high load condition. The statistical analysis was also performed for the observations where participants had four-out-of-four correct recalled digits in the high-load condition. Even under these more strenuous working memory load manipulations, the pattern of the results remained similar. The results of this exploratory analysis are presented in Appendix E.

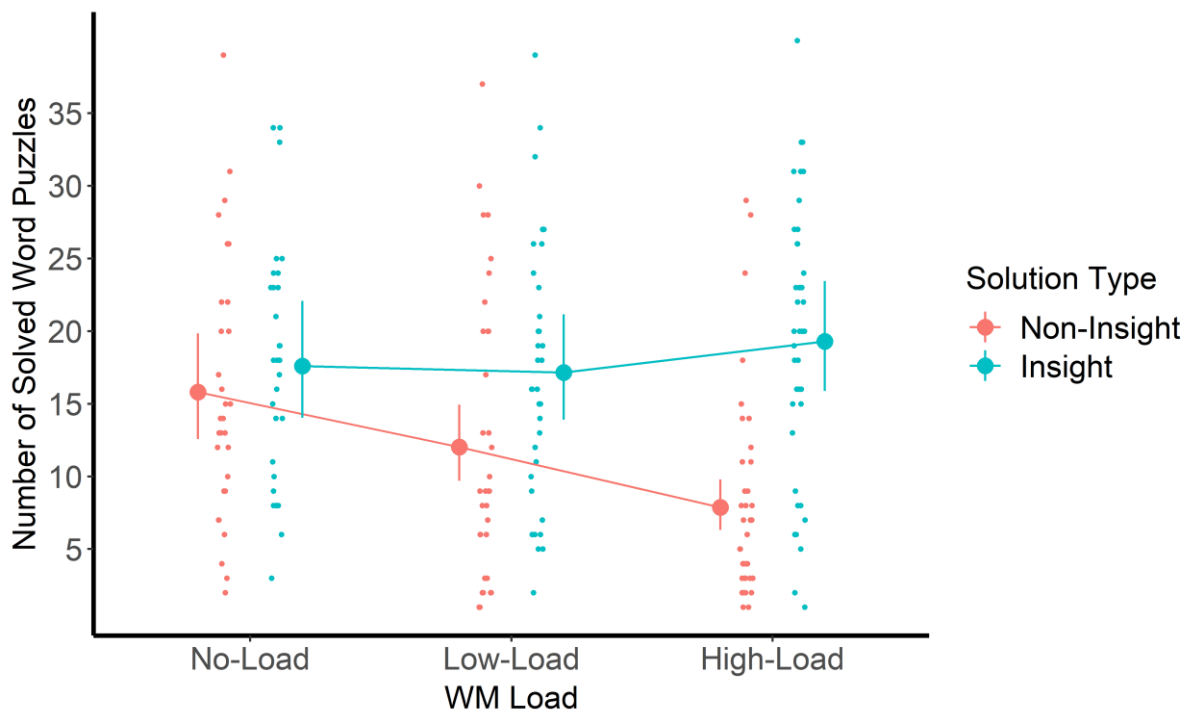


Figure 3. The interaction between solution type and WM load for the number of correctly solved word puzzles. Bars represent the 95% confidence intervals. The number of solved word puzzles refers to those solved correctly.

3.5 Secondary analysis

To assess whether the participant's self-reported insight solutions were independent of fast solutions, which were likely retrieved via immediate recognition, all estimated models described above were rerun, with the responses provided within the first two seconds excluded (see Appendix D). The pattern of results remained similar when compared to those obtained with the responses provided within the first two seconds included (see Appendix C). Namely, for solution time, a main effect of solution type ($\chi^2(1) = 14.54, p < .001$) and an interaction effect between solution type and WM load ($\chi^2(2) = 5.84, p = .054$) were observed. For solution accuracy and solution confidence only a main effect of solution type was found ($\chi^2(1) = 193.46, p < .001$ and $\chi^2(1) = 240.56, p < .001$, respectively). For the number of correctly solved word puzzles, main effects of solution type ($\chi^2(3) = 27.52, p < .001$) and WM load ($\chi^2(4) = 18.88, p < .001$) and an interaction effect between solution type and WM load ($\chi^2(2) = 10.08, p = .007$) were observed. All effects showed the same direction as the effects reported for the analyses, including the fast responses.

4 DISCUSSION

In the current study, we aimed to elucidate whether insight problem solving, as non-insight problem solving, relies on WM capacity (i.e., *Type 2 process*) or whether it is based on an unconscious process that does not tax cognitive resources (i.e., *Type 1 process*). To that end, we conducted a CRA experiment where participants solved word puzzles while concurrently executing a WM task.

Our results showed that the solution types were differentially influenced by limiting the available WM resources. Although correctly solved word puzzles with non-insight became less frequent with increasing WM loads, the number of correctly solved word puzzles with insight remained unaffected. Furthermore, in the high-load condition, solution time was longer for non-insight than insight solutions. These findings

support the dual-processing theory's topology for each solution type. Specifically, the observation that insight problem solving largely remained stable under dual-task conditions may indicate that insights' underlying processes operate outside awareness without taxing cognitive resources (*Type 1 process*). On the contrary, non-insight problem solving was impacted under dual-task conditions, which corroborates non-insights' dependence on WM resources (*Type 2 process*).

The exploratory analysis using a more extreme four-out-of-four correct criterium in the high-load condition further substantiated these findings. Even under these strenuous WM loads, the pattern of results remained similar to the initial analysis with a three-out-of-four correct criterium. This finding further strengthens the conception that insight depends on Type 1 processes more than non-insight does. Notwithstanding these findings, it is not unthinkable that even insight might have its limits as the intrinsic nature of problem solving entails some form of conscious processing at least. Recent insight theorists (Bowden & Grunewald, 2018; Weisberg, 2015) have argued that each solution type involves both processes (Type 1 and 2) albeit differentially so (see also Korovkin et al., 2018). Therefore, it would be interesting to identify under which exact circumstances insight can also be hampered. Of note, the exploratory analysis led to a large exclusion of data in the high-load condition. Although this confirms the extreme nature of the imposed load, the results should be approached cautiously given the high and unbalanced data exclusion.

To our knowledge, this is the first dual-task study on insight that addresses methodological issues of previous studies by only using one problem type (CRA) that can be solved both with insight and non-insight. Our results show that we succeeded in adequately capturing the two distinct solution types: insight solutions were more often correct, received higher solution confidence, and were solved more frequently than non-insightful ones. This closely mimics other findings typecasting insight problem solving (see Danek & Salvi, 2018; Danek & Wiley, 2017; Salvi et al., 2016; Webb et al., 2016). Moreover, we showed that using self-reports to classify word puzzles as solved with insight was not biased by fast responses via immediate recognition (see Cranford & Moss, 2012). Our secondary analysis where such fast responses were excluded displayed a similar pattern of results than the results where they were included. This finding shows that it was not just the fast responses that were identified as solved with insight. This is further supported by the data with the fast responses included, illustrating that in the no- and low-load conditions, there was no difference in the time needed to solve word puzzles with insight and non-insight. Therefore, we argue that our method provided a more reliable investigation of how insight and non-insight are (differentially) reliant on WM resources. This method should be used in future studies in more heterogeneous samples balanced for biological sex to increase the generalizability of the current findings. We observed some inter-individual differences in the propensity to solve the CRA puzzles either with insight or non-insight. Although most participants showed a balance between puzzles solved with insight and non-insight, some participants had a low rate of insight or non-insight solutions accompanied by a higher rate of the other solution type. Although we maximally accounted for these inter-individual differences in the estimated models, this observation poses an interesting avenue for further research to expose whether certain people indeed show a propensity to use a certain solution type. This would require a longitudinal approach where the robustness of the use of both strategies can be assessed across time.

Our findings align with several strands of research evidencing the unconscious nature of insight. For instance, studies using nearness-to-solution ratings while participants solve problems have shown that before finding a solution with insight, participants cannot indicate the nearness-to-solution, although, for non-insight, they can (Kizilirmak et al., 2018; Laukkonen et al., 2021; Metcalfe & Wiebe, 1987). Also, studies presenting subliminal solution cues during the solution search showed that this increased the insightful problem-solving performance, indicating that unconscious information affects insight (Bowden, 1997; Grant & Spivey, 2003; Hattori, Sloman, & Orita, 2013; also see Cristofori, Salvi, Jung-Beeman, & Grafman, 2018, for a subliminal reward priming study). Neuro-imaging and psychophysiological studies have likewise corroborated the unconscious nature of insight. For example, it has been found that there is a sudden burst

of gamma waves over the right temporal lobe at roughly the moment of insight. This burst of activation is assumably linked to the sudden convergence and integration of solution information as it surfaces consciousness (see Kounios & Jung-Beeman, 2014 for a review). On a similar note, Salvi et al. (2020) showed that immediately preceding solutions found with insight, but not with non-insight, there was a sudden increase in pupil dilation, marking the switch into awareness of the insightful solution.

However, insight might have been preceded by hunches, intuitions, or subtle experiences of cognitive conflict, steering the problem-solving process to the insight moment consciously (Schooler, Ohlsson, & Brooks, 1993; Schooler, 2011; Stuyck et al., 2021; Winkielman & Schooler, 2011; Zander et al., 2016). It has been argued that such ineffable phenomenological experiences are, like unconscious processes, relatively independent of cognitive resources (Schooler, 2011; Winkielman & Schooler, 2011). Therefore, our results might reflect such ineffable processing rather than true unconscious processing. However, what makes such experiences hard to put down in words (i.e., the ineffable part), is that they represent the felt component of tacit problem-related information present at the border of consciousness (Norman, Price & Duff, 2010; see also Bowers et al., 1990). Thus, even these ineffable processes, although having a link with a conscious experience, still represent implicit processing. In that sense, they do not necessarily negate insight's unconscious nature but perhaps instead refer to the involvement of unconscious processes to a different degree (i.e., gradual accumulation of intuitions leading to insight; Bowers et al., 1990, Zhang, Lei, & Li, 2016). Still, recent studies on nearness-to-solution ratings demonstrated a discontinuous pattern of these ratings for insight and a gradual one for non-insight, especially in the case of CRA (e.g., Kizilirmak et al., 2018; Laukkonen et al., 2021). If intuitions would play a prominent role in CRA insight solutions, one would also expect to find graduality in the nearness-to-solution ratings tapping into the insight. One could argue that the relatively short solving times observed for CRA puzzles prohibit the reliance on gradual intuitions. In more complex problem-solving settings, intermittently working on the problem might provide a better context for intuitions to develop as well as to act upon them (see Sio & Ormerod, 2009, for a review on incubation as a period where no attention is devoted to the problem).

Furthermore, if both solution types rely differentially on Type 1 and Type 2 processes (i.e., insight more on the former and non-insight more on the latter), it should be possible to doubly dissociate them. This would additionally involve assessing how insightful and non-insightful problem-solving performance are hampered differentially when a secondary task depletes the unconscious resources (i.e., Type 1). To our knowledge, only one attempt has been made so far to achieve this. Lebed and Korovkin (2017) used an implicit learning task (i.e., a classification task based on a hidden rule) as a secondary task next to a problem-solving task where participants solved an insight problem versus a non-insight problem. Their results showed that this secondary task only interfered with the insightful solution search. However, they did not assess the participant's awareness of the underlying principle of the implicit learning task. Therefore, it remains unclear whether the performance on the implicit learning task resulted from an unconscious or conscious comprehension of its underlying principles. Although it is an interesting study, we argue that it is exactly this awareness manipulation check of the secondary task that is crucial to be able to unambiguously dissociate the reliance of insight and non-insight on Type 1 and Type 2 processes, but which poses a challenge that is difficult to surmount (see Newell & Shanks, 2014 and Peters, Kentridge, Phillips, & Block, 2017, for a discussion). Moreover, they used two distinct problem types to tax insight and non-insight problem solving, making an unambiguous comparison difficult as these problem types are very different in nature. Nevertheless, a double dissociation is an exciting avenue to further corroborate that insight depends more on Type 1 processes than non-insight. However, such a study would entail theoretical and methodological pitfalls that are difficult to address.

Unexpectedly, solution accuracy and solution confidence of insight versus non-insight solutions were not differentially affected by WM load. Furthermore, the observed discrepancy regarding solution time only showed a trend towards an interaction effect and had a small effect size. One could argue that the WM load manipulation we used was insufficient to elicit stronger effects across all dependent measures. However, it

is noteworthy that even using a seemingly lenient three-out-of-four recalled digits as a criterium to indicate the accuracy of the WM task in the high load condition, participants experienced considerable difficulty recalling the digits. As a result, this three-out-of-four correctly recalled digits criterium already led to a substantial exclusion of the data (25%). This indicates that our WM load manipulation was sufficiently difficult. Still, we conducted an additional exploratory analysis (see Appendix E) using a four-out-of-four correctly recalled digits criterium to further increase the WM demands in the high load condition. Although this led to an additional reduction of the data of 25% in the high load condition (highlighting the difficulty of achieving this criterium), the pattern of results for solution accuracy and solution confidence remained similar. This indicates that solution accuracy and solution confidence, even under a more strenuous WM load condition, remained largely unaffected. The observation that solution accuracy remained unaffected under a more strenuous WM load condition strengthens the finding that the non-insight solution rate deteriorated as WM load increased. This is reflected in the fact that, in all analyses, non-insight solution accuracy was unaffected by WM load, whereas the number of correctly solved word puzzles with non-insight decreased over the WM load conditions. This implies that the amount of incorrectly solved word puzzles decreased proportionally to the amount of correctly solved word puzzles with increasing WM load. This provides further evidence that the overall rate of non-insight solutions dropped, regardless of how accurate they were.

Alternatively, the lack of a main effect of WM load on solution accuracy (see also Wieth & Burns, 2014) might not be due to a weak WM load manipulation, but rather to participants' strategy to safeguard performance effectiveness (i.e., solution accuracy) at the expense of processing efficiency (i.e., the effort needed to maintain a particular performance effectiveness, often defined as reaction or solution times; Derakshan & Eysenck, 2009). For insight, safeguarding performance effectiveness may be effortless. However, for non-insight, maintaining performance effectiveness under increased WM load might become challenging. As such, this increased effort to maintain a certain level of solution accuracy might lead to increased solution times under high-load, which indeed was the case in both the initial, secondary, and exploratory analyses.

In conclusion, dual-processing theories provide a useful framework to approach insight and non-insight problem-solving and their reliance on cognitive resources. We showed that insight's driving mechanisms are less dependent on WM resources than those of non-insight. Because we circumvented several methodological issues from previous dual-task studies on insight, this study provides more convincing evidence that the underlying processes of insight are more reminiscent of unconscious, Type 1 processes, whereas non-insight depends more on conscious, effortful Type 2 processes.

ACKNOWLEDGMENTS

We thank Laetitia Elewaut for the assistance with the data collection and the Fonds de la Recherche Scientifique for providing the opportunity to conduct this research under a research fellow grant.

FUNDING

This work was supported by the "Fonds de la Recherche Scientifique" [grant number 34736358, 2019].

Appendix A. Instructions given to participants about the solution types (translated from Dutch and Dutch example word puzzle replaced by an English puzzle)

Thank you for participating in this experiment.

During the experiment, you will be presented with three words. The goal is to find a word that you can attach to each of these three words so that three new meaningful words are created. For example: "cane/daddy /plum" is connected by the word "sugar", because with the word "sugar" the compound words "sugarcane/ sugar daddy/sugarplum" can be formed. For every word puzzle, the solution is always a word that you can only add either to the front or to the back of the three words.

Try to answer as quickly and accurately as possible. You have 25 s to find a solution. Once you have found the solution, press the space bar and enter your answer.

After you have solved a word puzzle, indicate your confidence in your solution. You can do this by using the cursor of the mouse to choose a position on a horizontal scale between "low confidence" and "high confidence".

Finally, you must indicate whether you have solved this word puzzle "with Aha!" or "without Aha!".

With Aha!: with an Aha! experience you become aware of the solution suddenly and clearly. This can be accompanied by a sense of revelation and relief.

Without Aha!: Unlike an Aha! feeling, finding a solution with analysis is characterized by a step-by-step search process.

Imagine a dark room that is suddenly lit up (with Aha!) or slowly lit with a dimmer switch (without Aha!). We ask you to indicate after each word puzzle if you have solved it "with Aha!" or "without Aha!".

Before you start solving the word puzzle, we will show you two numbers. The goal is that you try to retain these two numbers. The order is not important.

Immediately after that, you will have the opportunity to solve the word puzzle.

After you have solved the word puzzle, indicated your confidence in the solution, and whether you have solved this word puzzle "with Aha!" or "without Aha!", we will ask you to enter these two numbers.

Before the experiment starts, you can practice.

If something is still not clear, please call the experiment leader. Once all instructions are clear, press the spacebar to continue.

Of note, these instructions are an example of those presented in the low-load condition (i.e., with two digits). For the high-load condition, the number of digits to recall was four. In the no-load condition, no instructions were given with regards to the memory task.

Appendix B. Deviations from preregistration

Table 1 Deviations from preregistration (<https://osf.io/2v9yz>)

Preregistration	Adjustment
<p>Only participants were to be included who had an 80% accuracy on the WM tasks to ensure that participants jointly performed the CRA and WM tasks adequately.</p>	<p>Because this inclusion criterium led to a substantial omission of participants in the load-conditions (N=14 in the low-load and N=19 in the high-load condition), we adjusted this criterium so that we were able to include all participants, but only those CRA trials for which their WM task performance was correct.</p>
<p>We specified a secondary analysis to account for the influence of solutions found by immediate recognition on the pattern of results (see Cranford & Moss, 2012). To identify those types of solutions, we proposed a cut-off of 7470 ms.</p>	<p>The descriptive statistics of solution time demonstrated that the proposed cut-off was too lenient because the average CRA solution time was 7682 ms ($SD = 5035$). Hence, this proposed cut-off would not have reliably identified immediate recognitions and likely would also have included solutions found after an actual solution search. Therefore, we used a different exclusion criterium to account for the immediate recognition solutions by omitting all solutions found within the first 2000 ms (see Salvi et al., 2020).</p>
<p>An additional exploratory analysis was performed that was not included in the preregistration.</p>	<p>Initially, a three-out-of-four correctly recalled digits criterium was used to index accuracy on the WM task in the high-load condition. However, to assess the pattern of results under more strenuous WM demands, we increased this criterium in the high-load condition to a four-out-of-four correctly recalled digits criterium. Based on this altered criterium all statistical analyses were reconducted.</p>
<p>An additional exploratory analysis was performed that was not included in the preregistration.</p>	<p>The participant sample was not balanced for biological sex with 91 females and 14 males. Therefore, we reconducted the statistical analysis with biological sex as a covariate in each estimated model to assess the influence of biological sex on the obtained results.</p>
<p>An additional exploratory analysis was performed that was not included in the preregistration</p>	<p>The total number of correctly solved word puzzles varied among participants. To take this into account we reconducted the analysis of the number of correctly solved word puzzles with an offset included. Including this offset transforms the outcome variable to a rate (i.e., $\log(\text{number solved correctly}/\text{total number solved correctly})$) to estimate the model's parameters, after which the outcome variable is back-transformed to represent counts again (see Hutchinson & Holtman, 2005).</p>

Appendix C. Estimated models

Table 1. Linear mixed model on solution time of the CRA

	$\beta(SE)$	CI 95%	<i>t</i> -value	<i>p</i>
Intercept (grand mean)	3.84(0.01)	3.81, 3.87	269.78	<.001
solution type	0.04(0.01)	0.02, 0.06	3.91	<.001
Load1	-0.01(0.03)	-0.06, 0.04	-0.41	.686
Load2	0.007(0.03)	-0.04, 0.06	0.29	.773
solution type*load1	0.03(0.03)	-0.03, 0.09	1.05	.294
solution type*load2	0.07(0.03)	0.01, 0.12	2.41	.016

Note. Load1 = no vs. low; Load2 = low vs. high; *p*-values were obtained using the Satterthwaite approximation method; Boldface = significant results; CI = confidence interval. This linear mixed model is based on only the correctly solved word puzzles.

Table 2. Generalized linear mixed model on solution accuracy of the CRA

	$\beta(SE)$	OR	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	1.99(0.16)	7.32	5.34, 10.24	12.22	<.001
solution type	-1.56(0.12)	0.21	0.16, 0.26	-13.47	<.001
Load1	0.25(0.31)	1.28	0.69, 2.38	0.79	.429
Load2	-0.36(0.29)	0.70	0.39, 1.26	-1.21	.228
solution type*load1	-0.10(0.31)	0.90	0.49, 1.67	-0.33	.738
solution type*load2	-0.46(0.31)	0.63	0.34, 1.18	-1.46	.144

Note. Beta coefficients are on the logit scale; OR = odds ratio; An OR of one represents a chance-level classification of correct and incorrectly solved word puzzles. An OR above/below one represents the magnitude of increase/decrease in the chance of classifying word puzzles as solved correctly; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI = confidence interval. This generalized linear mixed model is based on the correct and incorrectly solved word puzzles.

Table 3. Generalized linear mixed model on solution confidence of the CRA

	$\beta(SE)$	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	0.97(0.09)	0.81, 1.14	11.31	<.001
solution type	-0.65(0.04)	-0.74, -0.57	-15.95	<.001
Load1	-0.22(0.18)	-0.58, 0.14	-1.18	.236
Load2	-0.39(0.17)	-0.72, -0.05	-2.24	.025
solution type*load1	-0.02(0.11)	-0.24, 0.20	-0.19	.852
solution type*load2	-0.11(0.12)	-0.34, 0.12	-0.95	.343

Note. Beta coefficients are on the logit scale; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI = confidence interval. This generalized linear mixed model is based on only the correctly solved word puzzles.

Table 4. Generalized Linear model on the number of correctly solved word puzzles of the CRA

	$\beta(SE)$	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	2.66(0.03)	2.60, 2.73	78.34	<.001
solution type	-0.45(0.11)	-0.66, -0.25	-4.30	<.001
Load1	-0.30(0.08)	-0.46, -0.14	-3.66	<.001
Load2	-0.30(0.11)	-0.51, -0.10	-2.85	.004
solution type*load1	-0.69(0.29)	-1.26, -0.13	-2.39	.017
solution type*load2	-0.89(0.30)	-1.48, -0.30	-2.96	.003

Note. Beta coefficients are on the log scale; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI, confidence interval. Robust standard errors were computed to account for the non-independence of the observations. This generalized linear model is only based on the correctly solved word puzzles.

Appendix D. Estimated models with responses within the two first seconds excluded

Table 1. Linear mixed model on solution time of the CRA

	$\beta(SE)$	CI 95%	<i>t</i> -value	<i>p</i>
Intercept (grand mean)	3.84(0.01)	3.82, 3.87	286.93	<.001
solution type	0.04(0.01)	0.02, 0.06	3.82	<.001
Load1	-0.01(0.03)	-0.06, 0.05	-0.27	.785
Load2	0.003(0.02)	-0.05, 0.05	0.12	.909
solution type*load1	0.02(0.03)	-0.04, 0.08	0.80	.422
solution type*load2	0.07(0.03)	0.01, 0.13	2.38	.018

Note. Load1 = no vs. low; Load2 = low vs. high; *p*-values were obtained using the Satterthwaite approximation method; Boldface = significant results; CI = confidence interval. This linear mixed model is based on only the correctly solved word puzzles.

Table 2. Generalized Linear mixed model on solution accuracy of the CRA

	$\beta(SE)$	OR	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	1.99(0.16)	7.32	5.34, 10.26	12.20	<.001
solution type	-1.56(0.12)	0.21	0.17, 0.26	-13.35	<.001
Load1	0.25(0.31)	1.28	0.69, 2.39	0.80	.423
Load2	-0.34(0.29)	0.71	0.39, 1.28	-1.17	.243
solution type*load1	-0.10(0.31)	0.91	0.49, 1.68	-0.31	.756
solution type*load2	-0.50(0.31)	0.61	0.32, 1.13	-1.61	.108

Note. Beta coefficients are on the logit scale; OR = odds ratio; An OR of one represents a chance-level classification of correct and incorrectly solved word puzzles. An OR above/below one represents the magnitude of increase/decrease in the chance of classifying word puzzles as solved correctly; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI = confidence interval. This generalized linear mixed model is based on the correct and incorrectly solved word puzzles.

Table 3. Generalized Linear mixed model on solution confidence of the CRA

	$\beta(SE)$	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	0.97(0.09)	0.80, 1.14	11.26	<.001
solution type	-0.65(0.04)	-0.74, -0.57	-15.84	<.001
Load1	-0.22(0.18)	-0.58, 0.14	-1.18	.238
Load2	-0.39(0.17)	-0.72, -0.05	-2.23	.026
solution type*load1	0.0004(0.11)	-0.22, 0.22	0.004	.997
solution type*load2	-0.10(0.12)	-0.33, 0.13	-0.88	.378

Note. Beta coefficients are on the logit scale; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI = confidence interval. This generalized linear mixed model is based on only the correctly solved word puzzles.

Table 4. Generalized Linear model on the number of correctly solved word puzzles of the CRA

	$\beta(SE)$	CI 95%	<i>Z</i> -value	<i>p</i>
Intercept (grand mean)	2.65(0.03)	2.58, 2.71	78.03	<.001
solution type	-0.45(0.10)	-0.65, -0.24	-4.30	<.001
Load1	-0.31(0.08)	-0.47, -0.15	-3.80	<.001
Load2	-0.30(0.11)	-0.50, -0.09	-2.79	.005
solution type*load1	-0.68(0.29)	-1.24, -0.12	-2.37	.018
solution type*load2	-0.88(0.30)	-1.46, -0.30	-2.97	.003

Note. Beta coefficients are on the log scale; Load1 = no vs. low; Load2 = low vs. high; Boldface = significant results; CI, confidence interval. Robust standard errors were computed to account for the non-independence of the observations. This generalized linear model is based on only the correctly solved word puzzles.

Appendix E. The statistical analysis with the four-out-of-four correctly recalled digits criterium

Solution time

An LMM was constructed with the (log-transformed) solution time of the correctly solved word puzzles as the outcome variable and solution type (two levels: Insight and non-insight), WM load (three levels: no-load, low-load, and high-load) and their interaction term as fixed effects. There was a significant main effect of solution type, $\chi^2(1) = 14.82$, $p < .001$. Word puzzles solved with insight ($M = 3.81$) were solved significantly faster than those solved with non-insight ($M = 3.86$), $t(2684) = -3.85$, $p < .001$, Cohen's $d = 0.18$ (95% CI [0.09, 0.27]). There also was a significant interaction effect, $\chi^2(2) = 5.99$, $p = .049$. Post-hoc tests showed that there were no significant differences between load conditions for insight (no load vs. low load, $p = .878$; no load vs. high load, $p = .338$; and low load vs. high load, $p = .609$) nor non-insight (no load vs. low load, $p = .793$; no load vs. high load, $p = .748$; and low load vs. high load, $p = .382$). To further clarify the significant interaction effect, a pairwise contrast was made of solution type (insight vs. non-insight) conditional on three levels of WM load. These pairwise comparisons showed that only in the high-load condition, word puzzles solved with insight were solved significantly faster than word puzzles solved with non-insight ($M = 3.79$ versus $M = 3.88$), $t(2616) = -3.83$, $p < .001$, Cohen's $d = 0.34$ (95% CI [0.17, 0.52]). This was not the case in the no- and low-load conditions ($p = .146$ and $p = .265$). The main effect of WM load was not significant, $p = .809$. This result is depicted in Figure 1.

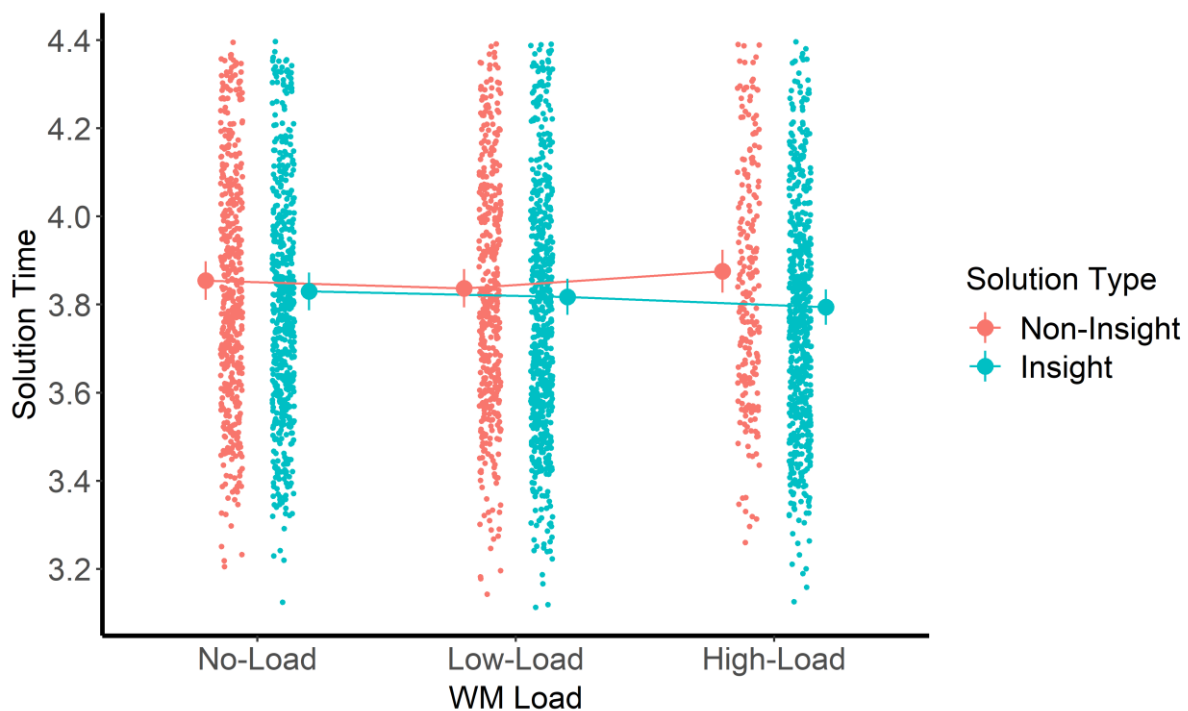


Figure 1. The interaction between solution type and WM load for solution time. Bars represent the 95% confidence intervals. Solution time was log-transformed with base 10.

Solution accuracy

Likewise, a GLMM was constructed with solution accuracy as a binary outcome variable. There was a significant main effect of solution type, $\chi^2(1) = 161.9$, $p < .001$. Word puzzles solved with insight had a significantly higher probability of being correct than word puzzles solved with non-insight ($M = 94\%$ versus $M = 78\%$), $Z = 12.28$, $p < .001$, Cohen's $d = -1.52$ (95% CI [-1.76, -1.27]). The main effect of WM load and the interaction effect were not significant, $p = .184$ and $p = .724$, respectively.

Solution confidence

Similarly, a GLMM was constructed with solution confidence of the correctly solved word puzzles as a bounded outcome variable (range .005-.995). There was a significant main effect of solution type, $\chi^2(1) = 194.63$, $p < .001$. The word puzzles solved with insight received significantly higher solution confidence than word puzzles solved with non-insight ($M = .78$ versus $M = .66$), $t(2825) = 14.23$, $p < .001$, Cohen's $d = -0.14$ (95% CI [-0.16, -0.12]). The main effect of WM load and the interaction effect were not significant, $p = .125$ and $p = .927$, respectively.

The number of correctly solved word puzzles

Lastly, a GLM was built with the number of correctly solved word puzzles as a count outcome variable⁴. There was a significant main effect of solution type, $\chi^2(3) = 23.73$, $p < .001$. Correctly solved word puzzles were more frequent for insight than non-insight ($M = 16.4$ versus $M = 10.5$), $Z = 4.89$, $p < .001$, Cohen's $d = -0.43$ (95% CI [-0.61, -0.25]). Also, a main effect of WM load was observed, $\chi^2(4) = 36.64$, $p < .001$. Post-hoc tests demonstrated that correctly solved word puzzles were more frequent in the no- than high-load condition ($M = 16.7$ versus $M = 9.4$), $Z = -5.00$, $p < .001$, Cohen's $d = 0.55$ (95% CI [0.33, 0.77]) and more frequent in the low- than high load condition ($M = 14.4$ versus $M = 9.4$), $Z = -3.82$, $p < .001$, Cohen's $d = 0.40$ (95% CI [0.19, 0.61]). The comparison between the no- and low-load condition was not significant ($p = .389$). More importantly, there was a significant interaction effect, $\chi^2(2) = 9.03$, $p = .011$. The post-hoc tests illustrated that for insight, there were no significant differences between the three levels of WM load (no-load vs. low-load, $p = .985$; no-load vs. high-load, $p = .480$ and low-load vs. high load, $p = .560$). However, for non-insight correctly solved word puzzles were more frequent in the no- than high-load condition ($M = 15.8$ versus $M = 6.0$, $Z = -5.76$, $p < .001$, Cohen's $d = 0.92$ (95% CI [0.59, 1.25]), and more frequent in the low- than high-load condition ($M = 12.0$ versus $M = 6.0$, $Z = -4.24$, $p < .001$, Cohen's $d = 0.66$ (95% CI [0.35, 0.97]). The comparison between the no- and low-load condition was not significant ($p = .214$). Thus, while the number of correctly solved word puzzles decreased with increasing WM loads for non-insight, the number of correctly solved word puzzles with insight remained unaffected. This result is illustrated in Figure 2.

⁴ As the total number of correctly solved word puzzles varied among participants, we took this into account by including an offset in the GLM of the number of correctly solved word puzzles. Including this offset transforms the outcome variable to a rate (i.e., $\log(\text{number solved correctly}/\text{total number solved correctly})$) to estimate the model's parameters (see Hutchinson & Holtman, 2005). The results of this adjusted GLM remained similar to the initial analysis. Namely, a main effect of solution type ($p < .001$), a main effect of WM load ($p < .001$), and an interaction effect between solution type and WM load ($p = .008$). The direction of the effects remained similar to the initial analysis.

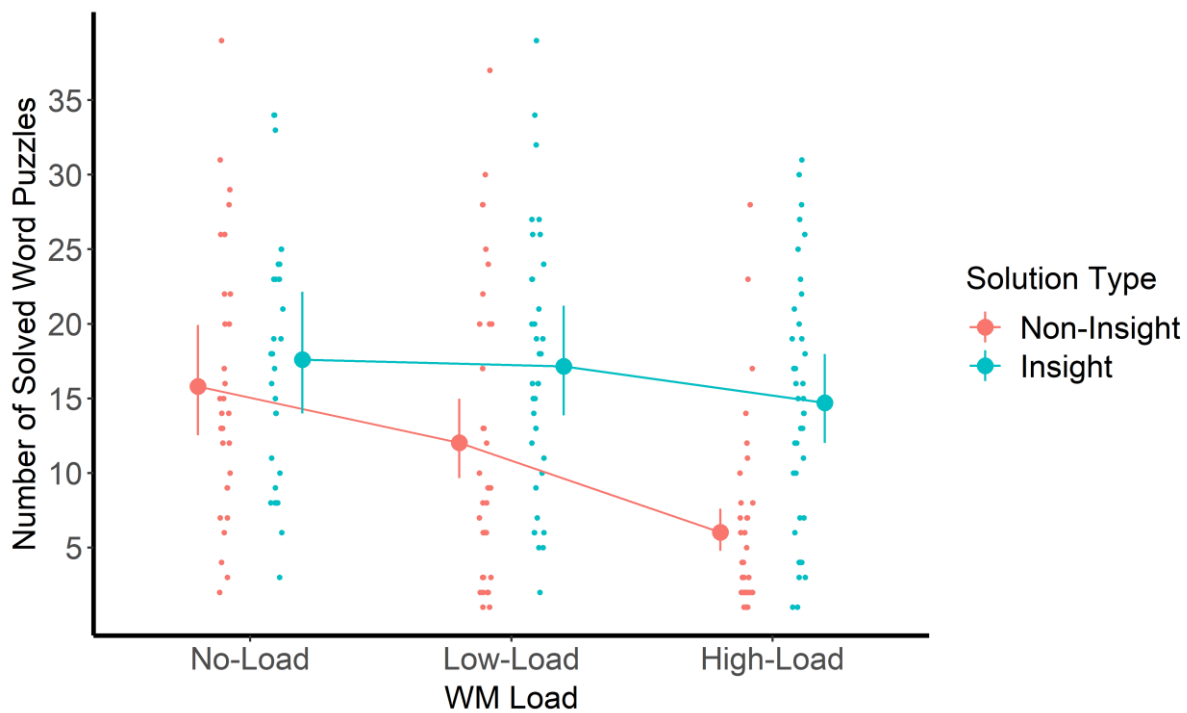


Figure 2. The interaction between solution type and WM load for the number of correctly solved word puzzles. Bars represent the 95% confidence intervals. The number of solved word puzzles refers to those solved correctly.

Stuyck, H., Cleeremans, A., & Van den Bussche, E. (2021). Aha! under pressure: The Aha! experience is not constrained by cognitive load. *Open Science Framework (OSF)*.
<https://doi.org/10.17605/OSF.IO/SC5N7>

References

- Abadie, M., Waroquier, L., & Terrier, P. (2013). Gist memory in the unconscious-thought effect. *Psychological Science, 24*(7), 1253–1259. <https://doi.org/10.1177/0956797612470958>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.
<https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28. <https://doi.org/10.21500/20112084.807>
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Bago, B., & De Neys, W. (2019). The smart system 1: evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking and Reasoning, 25*(3), 257–299.
<https://doi.org/10.1080/13546783.2018.1507949>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models.
<https://arxiv.org/abs/1506.04967>.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–29. <https://doi.org/10.18637/jss.v067.i01>
- Benedek, M., & Fink, A. (2019). Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control. *Current Opinion in Behavioral Sciences, 2019*(27), 116–122.
<https://doi.org/10.1016/j.cobeha.2018.11.002>
- Bowden, E. M. (1997). The effect of reportable and unreportable hints on anagram solution and the Aha! experience. *Consciousness and Cognition, 6*(6), 545–573. <https://doi.org/10.1006/ccog.1997.0325>
- Bowden, E. M., & Grunewald, K. (2018). Whose insight is it anyway? In F. Vallée-Tourangeau (Ed.), *Insight: On the Orgins of New Ideas* (pp. 28–50). New York: Routledge.
- Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35*(4), 634–639.
<https://doi.org/10.3758/BF03195543>
- Bowden, E. M., & Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods, 42*(2007), 87–99. <https://doi.org/10.1016/j.ymeth.2006.11.007>
- Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9*(7), 322–328. <https://doi.org/10.1016/j.tics.2005.05.012>
- Bowers, J., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the Context of Discovery, *10*, 72–110.

- Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), 378–400.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. <https://doi.org/10.5334/joc.10>
- Camarda, A., Borst, G., Agogu e, M., Habib, M., Weil, B., Houd e, O., & Cassotti, M. (2018). Do we need inhibitory control to be creative? Evidence from a dual-task paradigm. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 351–358. <https://doi.org/10.1037/aca0000140>
- Chater, N. (2018). *The Mind is Flat: The Illusion of Mental Depth and the Improvised Mind*. UK: Penguin Books LTD.
- Chein, J. M., & Weisberg, R. W. (2014). Working memory and insight in verbal problems: Analysis of compound remote associates. *Memory and Cognition*, *42*(1), 67–83. <https://doi.org/10.3758/s13421-013-0343-4>
- Chuderski, A., & Jastrzebski, J. (2018). Much ado about aha!: Insight problem solving is strongly related to working memory capacity and reasoning ability. *Journal of Experimental Psychology: General*, *147*(2), 257–281. <https://doi.org/10.1037/xge0000378>
- Cooper, R. P., & Marsh, V. (2015). Set-shifting as a component process of goal-directed problem-solving. *Psychological Research*, *80*(2), 307–323. <https://doi.org/10.1007/s00426-015-0652-2>
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- Cranford, E. A., & Moss, J. (2012). Is insight always the same? A protocol analysis of insight in compound remote associate problems. *The Journal of Problem Solving*, *4*(2), 128–153. <https://doi.org/10.7771/1932-6246.1129>
- Cristofori, I., Salvi, C., Jung-Beeman, M., & Grafman, J. (2018). The effects of expected reward on creative problem solving. *Cognitive, Affective and Behavioral Neuroscience*, *18*(5), 925–931. <https://doi.org/10.3758/s13415-018-0613-5>
- Danek, A. H., Fraps, T., von M uller, A., Grothe, B., &  ollinger, M. (2014). Working wonders? Investigating insight with magic tricks. *Cognition*, *130*(2), 174–185. <https://doi.org/10.1016/j.cognition.2013.11.003>
- 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. <https://doi.org/10.3389/fpsyg.2016.02077>
- Danek, A. H., Wiley, J., &  ollinger, M. (2016). Classical insight problems without Aha! experience : 9 Dot , 8 Coin , and matchstick arithmetic problems. *Journal of Problem Solving*, *9*, 47–57. <https://doi.org/10.7771/1932-6246.1183>.
- Danek, A., & Salvi, C. (2018). Moment of truth : Why Aha ! experiences are correct. *Journal of Creative Behavior*, *0*, 1–3. <https://doi.org/10.1002/jocb.380>
- De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–434. doi: 10.1111/j.1467-9280.2006.01723.x. PMID: 16683931.

- De Neys, W. (2021). On dual- and single-process models of thinking. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691620964172>
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503–509. <https://doi.org/10.1177/0963721419855658>
- DeCaro, M. S., Van Stockum, C. A., & Wieth, M. B. (2016). When higher working memory capacity hinders insight. *Journal of Experimental Psychology: Learning Memory and Cognition*, 42(1), 39–49. <https://doi.org/10.1037/xlm0000152>
- Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *European Psychologist*, 14(2), 168–176. <https://doi.org/10.1027/1016-9040.14.2.168>
- De Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, 38(5), 656–669. <https://doi.org/10.1177/0146167211435795>
- Evans, J. S. B. T. (2019). Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning. *Thinking and Reasoning*, 25(4), 383–415. <https://doi.org/10.1080/13546783.2019.1623071>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *European Journal of Cognitive Psychology*, 20(1), 139–176. <https://doi.org/10.1080/09541440601016954>
- Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, 25(4), 436–463. <https://doi.org/10.1080/20445911.2013.779248>
- Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology (2006)*, 71(5), 1188–1208. <https://doi.org/10.1080/17470218.2017.1313283>
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392–404. <https://doi.org/10.1037/0033-2909.118.3.392>
- Gilhooly, K. J., Ball, L. J., & Macchi, L. (2015). Insight and creative thinking processes: Routine and special. *Thinking and Reasoning*, 21(1), 1–4. <https://doi.org/10.1080/13546783.2014.966758>
- Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking & Reasoning*, 15(4), 355–376. <https://doi.org/10.1080/13546780903178615>
- 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>
- 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. <https://doi.org/10.3389/fpsyg.2016.01314>
- Howarth, S., Handley, S. J., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict. *Memory and Cognition, 44*(2), 330–349. <https://doi.org/10.3758/s13421-015-0555-x>
- Hutchinson, M. K., & Holtman, M. C. (2005). Analysis of count data using poisson regression. *Research in Nursing and Health, 28*(5), 408–18. doi: 10.1002/nur.20093.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology : A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. <https://doi.org/10.1037/a0028347>
- Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-liu, S., Greenblatt, R., ... Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology, 2*(4), 500–510. <https://doi.org/10.1371/journal.pbio.0020097>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kizilirmak, J. M., Serger, V., Kehl, J., Öllinger, M., Folta-Schoofs, K., & Richardson-Klavehn, A. (2018). Feelings-of-warmth increase more abruptly for verbal riddles solved with in contrast to without aha! Experience. *Frontiers in Psychology, 9*, 1440. <https://doi.org/10.3389/fpsyg.2018.01404>
- Korovkin, S., Vladimirov, I., Chistopolskaya, A., & Savinova, A. (2018). How working memory provides representational change during insight problem solving. *Frontiers in Psychology, 9*, 1864. <https://doi.org/10.3389/fpsyg.2018.01864>
- Korovkin, S. Y., & Savinova, A. D. (2014). The dynamics of working memory load in insight problem solving. *The Russian Journal of Cognitive Science, 1*(4), 67–81.
- Kounios, J., & Jung-Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology, 65*, 71–87. <https://doi.org/10.1146/annurev-psych-010213-115154>
- 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, 0*(0), 1–18. <https://doi.org/10.1080/02699931.2021.1908230>
- 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*. <https://doi.org/10.1016/j.cognition.2019.104122>
- Laukkonen, R., Webb, M., Salvi, C., Tangen, J. M., & Schooler, J. (2018). Eureka heuristics: How feelings of insight signal the quality of a new idea. <https://doi.org/10.31234/osf.io/ez3tn>
- Lavric, A., Forstmeier, S., & Rippon, G. (2000). Differences in working memory involvement in analytical and creative tasks : an ERP study. *Neuroreport, 11*(8):1613-1618. doi: 10.1097/00001756-200006050-00004
- Lebed, A. A., & Korovkin, S. Y. (2017). The unconscious nature of insight : A dual-task paradigm investigation. *Psychology in Russia: State of the Art, 10*(3). <https://doi.org/10.11621/pir.2017.0307>
- Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-square means. R package version 1.4.7. <https://CRAN.R-project.org/package=emmeans>

- Levy, R. (2018). Using R formulae to test for main effects in the presence of higher-order interactions, 1–7. Retrieved from <http://arxiv.org/abs/1405.2094>
- Lin, W. L., & Lien, Y. W. (2013). The different role of working memory in open-ended versus closed-ended creative problem solving: A dual-process theory account. *Creativity Research Journal*, *25*(1), 85–96. <https://doi.org/10.1080/10400419.2013.752249>
- Mega, L. F., & Volz, K. G. (2014). Thinking about thinking: Implications of the introspective error for default-interventionist type models of dual processes. *Frontiers in Psychology*, *5*, 864. <https://doi.org/10.3389/fpsyg.2014.00864>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–15. doi: 10.1016/j.jml.2017.01.001.
- Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*(3), 238–246. <https://doi.org/10.3758/BF03197722>
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *The Behavioral and Brain Sciences*, *37*(1), 1–19. <https://doi.org/10.1017/S0140525X12003214>
- Norman, E., Price, M. C., & Duff, S. C. (2010). Fringe Consciousness: A Useful Framework for Clarifying the Nature of Experience-Based Metacognitive Feelings. In A. Efklides & M. Plousia (Eds.), *Trends and Prospects in Metacognition Research*. New York: Springer.
- Oberauer, K., (2019). Working memory and attention- A conceptual analysis and review. *Journal of Cognition* 2(1): 36, 1-23. <https://doi.org/10.5334/joc.58>
- Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What Limits Working Memory Capacity? *Psychological Bulletin*, *142*(7), 758–799.
- Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). Harvester/Wheatsheaf.
- Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. New York: Cambridge University Press.
- Öllinger, M., & von Müller, A. (2017). Search and coherence-building in intuition and insight problem Solving. *Frontiers in Psychology*, *8*, 827. <https://doi.org/10.3389/fpsyg.2017.00827>
- Peters, M. A. K., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, *2017*(1), 1–11. <https://doi.org/10.1093/nc/nix015>
- Pliatsikas, C., Verissimo, J., Babcock, L., Pullman, M. Y., Gleib, D. A., Weinstein, M., ... Ullman, M. T. (2019). Working memory in older adults declines with age, but is modulated by sex and education. *Quarterly Journal of Experimental Psychology*, *72*(6), 1308–1327. <https://doi.org/10.1177/1747021818791994>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management*, *36*(4), 941–973. <https://doi.org/10.1177/0149206309350084>

- Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Jung-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>
- Salvi, C., Simoncini, C., Grafman, J., & Jung-Beeman, M. (2020). Oculometric signature of switch into awareness? Pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis. *NeuroImage*, *217*(September 2019), 116933.
<https://doi.org/10.1016/j.neuroimage.2020.116933>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*(October 2019), 104038.
<https://doi.org/10.1016/j.jml.2019.104038>
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime Users' Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schooler, Jonathan W; Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*(2), 166–183.
- Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, *137*(2), 345–350. <https://doi.org/10.1037/a0022390>
- Shen, W., Tong, Y., Yuan, Y., Zhan, H., Liu, C., Luo, J., & Cai, H. (2017). Feeling the insight: uncovering somatic markers of the "aha" Experience. *Applied Psychophysiology Biofeedback*, *43*, 13–21.
<https://doi.org/10.1007/s10484-017-9381-1>
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: maintenance and disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799.
<https://doi.org/10.1177/1745691616650647>
- Simon, H. A., & Newell, A. (1970). Human problem solving: The state of theory in 1970. *American Psychologist*, *26*(2), 145-150. <https://doi.org/10.1037/h0030806>
- 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>
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71.
<https://doi.org/10.1037/1082-989X.11.1.54>
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling : A simplified three-step procedure using Stata , R , Mplus , and SPSS. *International Review of Social Psychology*, *30*(1), 203–218.
- Sowden, P. T., Pringle, A., & Gabora, L. (2015). The shifting sands of creative thinking: Connections to dual-process theory. *Thinking and Reasoning*, *21*(1), 40–60.
<https://doi.org/10.1080/13546783.2014.885464>
- Stuyck, H., Aben, B., Cleeremans, A., & Van den Bussche, E. (2021). The Aha! moment: Is insight a different form of problem solving. *Consciousness and Cognition*, *90*.
<https://doi.org/10.1016/j.concog.2020.103055>
- Venables, W. N., and B. D. Ripley. (2002). *Modern Applied Statistics with S* (4th ed.). New York: Springer.

- Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, 37(1), 82–113. <https://doi.org/10.3102/1076998610396895>
- 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. <https://doi.org/10.3389/fpsyg.2016.01424>
- Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, 21(1), 5–39.
- 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). New York: Routledge.
- Wieth, M. B., & Burns, B. D. (2014). Rewarding multitasking: Negative effects of an incentive on problem solving under divided attention. *Journal of Problem Solving*, 7(1), 60–72. <https://doi.org/10.7771/1932-6246.1163>
- Wieth, M., & Burns, B. D. (2006). Incentives improve performance on both incremental and insight problem solving. *Quarterly Journal of Experimental Psychology*, 59(8), 1378–1394. <https://doi.org/10.1080/17470210500234026>
- Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 56. The psychology of learning and motivation* (p. 185–227). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-394393-4.00006-6>
- Winkielman, P., & Schooler, J. W. (2011). Splitting consciousness: Unconscious, conscious, and metaconscious processes in social cognition. *European Review of Social Psychology*, 22(1), 1–35. <https://doi.org/10.1080/10463283.2011.576580>
- Xu, X. S., Samtani, M., Yuan, M., & Nandy, P. (2014). Modeling of bounded outcome scores with data on the boundaries: Application to disability assessment for dementia scores in alzheimer's disease. *AAPS Journal*, 16(6), 1271–1281. <https://doi.org/10.1208/s12248-014-9655-y>
- Zander, T., Öllinger, M., & Volz, K. G. (2016). Intuition and insight: Two processes that build on each other or fundamentally differ? *Frontiers in Psychology*, 7, 1395. <https://doi.org/10.3389/fpsyg.2016.01395>
- Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News* 2(3), 7–10.
- Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object oriented implementation of clustered covariances in r." *Journal of Statistical Software* 95:1–36. doi: 10.18637/jss.v095.i01.
- Zhang, Z., Lei, Y., & Li, H. (2016). Approaching the distinction between intuition and insight. *Frontiers in Psychology*, 7, 1195. <https://doi.org/10.3389/fpsyg.2016.01195>