Insight and non-insight problem solving: A heart rate variability study

Stuyck Hans[1, 2]*, Demeyer Febe[1], Bratanov, Christo[1], Cleeremans Axel [2] & Van den Bussche Eva[1]

[1]*KU Leuven, Faculty of Psychology and Educational Sciences, Brain & 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
febe.demeyer@kuleuven.be
chr.bratanov@gmail.com
axcleer@ulb.ac.be
eva.vandenbussche@kuleuven.be

ABSTRACT

Occasionally, problems are solved with a sudden Aha! Moment (insight), while the mundane approach to solve problems is analytical (non-insight). At first glance, non-insight appears to depend on the availability and taxation of cognitive resources to execute the step-by-step approach, whereas insight does not or to a lesser extent. However, this remains debated. To investigate the reliance of both solution types on cognitive resources, we assessed the involvement of the prefrontal cortex using vagally mediated heart rate variability (vmHRV) as an index. Participants (N = 68) solved 70 compound remote associates word puzzles solvable with insight and non-insight. Before, during, and after solving the word puzzles, we measured vmHRV. Our results showed that resting-state vmHRV (trait) showed a negative association with behavioural performance for both solution types. This might reflect inter-individual differences in inhibitory control. As the solution search requires one to think of remote associations, inhibitory control might hamper rather than aid this process. Furthermore, we observed, for both solution types, a vmHRV increase from resting-state to solution search (state), lingering on in the post-task recovery period. This could mark the increase of prefrontal resources to promote an open-minded stance, essential for divergent thinking, which arguably is crucial for this task. Our findings suggest that, at a general level, both solution types share common aspects. However, a closer analysis of early and late solutions and puzzle difficulty suggested that metacognitive differentiation between insight and non-insight improved with higher trait vmHRV, and that a unique association between trait vmHRV and puzzle difficulty was present for each solution type.

Keywords: Aha! experience, insight problem solving, heart rate variability, prefrontal cortex

1. INTRODUCTION

As people go about their daily lives, they are often confronted with problems requesting their undivided attention, such as changing a flat tire or tackling monthly finances. Most of these problems can be solved by relying on stored knowledge and procedures, such as factual knowledge, heuristics, and prior problem solving experiences accumulated through life, for instance, via education (Weisberg, 2015). One of the prominent features of this analytical problem-solving strategy is that it is effortful and demands attention (Stuyck et al., 2022). It requires the problem solver to maintain a representation of the problem while applying different solution strategies and to avoid being distracted by irrelevant information (Shipstead et al., 2016; Wiley & Jarosz, 2012). Although analytical problem-solving is the general modus operandi to solve complex problems, occasionally, a solution is also obtained by a sudden epiphany. The felt component of such an epiphany is known as the Aha! experience or Eureka moment, and we will refer to this phenomenon as insight (Danek et al., 2014). At first glance, one of insight's defining features that sets it apart from the analytical, non-insight strategy is the apparent effortlessness of solution retrieval: No specific overt problem-solving behaviour seems to precede the insightful solution (e.g., Laukkonen et al., 2021). This conception of insight fits well with the phenomenological reports of different great thinkers, such as Judah Folkman and John Nash (Kounios & Jung-Beeman, 2015; Laukkonen et al., 2020). The great French physicist André-Marie Ampère described his experience of insight as follows:

> *"I gave a shout of joy. ... It was seven years ago I proposed to myself a problem which I have not been able to solve directly, but for which I had found by chance a solution, and knew that it was correct, without being able to prove it. The matter often returned to my mind and I had sought twenty times unsuccessfully for this solution. For some days I had carried the idea about with me continually. At last, I do not know how, I found it, together with a large number of curious and new considerations concerning the theory of probability."* (Horvitz, 2002, p. 1)

This description vividly illustrates the out-of-the-blue, effortless nature of insight. Such phenomenological reports have spurred some insight theorists to propose that insight mainly arises via unconscious processes (Fedor et al., 2017; Jung-Beeman et al., 2004). For example, it has been argued that prior assumptions and knowledge misguide the problem solver to an over-constrained solution space. As a result, the problem solver cannot find the solution and reaches a dead end. This state of impasse subsequently propagates negative feedback through the information processing system, thereby decreasing the activation of the misguiding assumptions and redistributing the unconscious spreading of activation to more remotely related concepts needed to relax the self-imposed constraints. As those constraints are relaxed, the whole solution space is restructured, eventually revealing the correct solution path (Fedor et al., 2017; Ohlsson, 2011, 2018). Other theorists, however, have questioned this conceptualization of insight (Benedek & Fink, 2019; Weisberg, 2018) and have argued that insight and non-insight are more alike than meets the eye. In their view, restructuring the initial erroneous solution space is achieved by consciously and incrementally building on new information arising from failed solution attempts (MacGregor et al., 2001; Weisberg, 2015). Proponents of each theory have suggested either a mainly unconscious or conscious trajectory leading to an insightful solution, although they concede that both insight and non-insight rely on conscious and unconscious processes to attain the solution, albeit to differing degrees (Becker et al., 2021; Bowden & Grunewald, 2018; Weisberg, 2015). It seems clear that non-insight requires a conscious mental workspace where information can be manipulated to reach task goals. Working memory (WM) is often referred to as such a workspace (e.g., Frith, 2021, Wiley & Jarosz, 2012). However, it remains unclear whether insight relates similarly to this conscious WM workspace. In the following, we briefly overview the relationship between consciousness, working memory, and the two solution types (i.e., insight and non-insight).

1.1 Consciousness, working memory, and the two solution types

It has been argued that the difference between conscious and unconscious states is that conscious states result from large-scale brain connections, with sensory cortices providing content-specific information and prefrontal and parietal cortices collating this information into rich subjective experiences (see, for a discussion, Frith, 2021). Hence, according to such perspectives, conscious access to mental representations depends on the prefrontal and parietal cortices (Dehaene, 2014), precisely those areas also involved in the goal-directed manipulation of mental representations (i.e., cognitive control; Martin-Signes et al., 2020; Riddle et al., 2020). Trübutschek et al. (2019) illustrated the interwoven character of consciousness and cognitive control by showing that cognitive manipulation of mental representations in service of task goals requires consciousness (but see also Sklar et al., 2021, for an opposing view). One central prefrontal processing hub where information is consciously manipulated to cope with ongoing task demands is WM (Chuderski & Jastrzebski, 2018; Frith, 2021; Funahashi, 2017). This hub enables people to maintain the problem representation, to focus their attention on successful instead of unsuccessful solution paths, to switch between solution strategies, and to steer away from irrelevant information (Shipstead et al., 2016; Wiley & Jarosz, 2012). The capacity of WM (WMC), which one may think is related to problem solving ability, has been approached from a trait perspective (i.e., inter-individual differences) and a resource-dependent perspective (i.e., limited cognitive resources). From a *trait perspective*, it is argued that people with a higher WMC than those with a lower WMC can sustain more task-relevant information in WM. This enables them to solve problems more efficiently than low-WMC people because they can simultaneously consider a broader spectrum of solution approaches to navigate the solution space (Unsworth et al., 2014). For example, it has been found that people's WMC is positively associated with more effective strategy use and performance in complex problem-solving (e.g., Ellis & Brewer, 2018; Gonthier & Roulin, 2020; Peng et al., 2016). From a *resource perspective*, WMC is conceived as a limited cognitive resource that can become depleted when increasing amounts of information require attention to cope with one or more ongoing tasks (Oberauer, 2019; Oberauer et al., 2016). For example, a recent meta-analysis showed that performing a secondary task while solving arithmetic problems (i.e., dual-task paradigm) hampered arithmetic performance (Chen and Bailey, 2021).

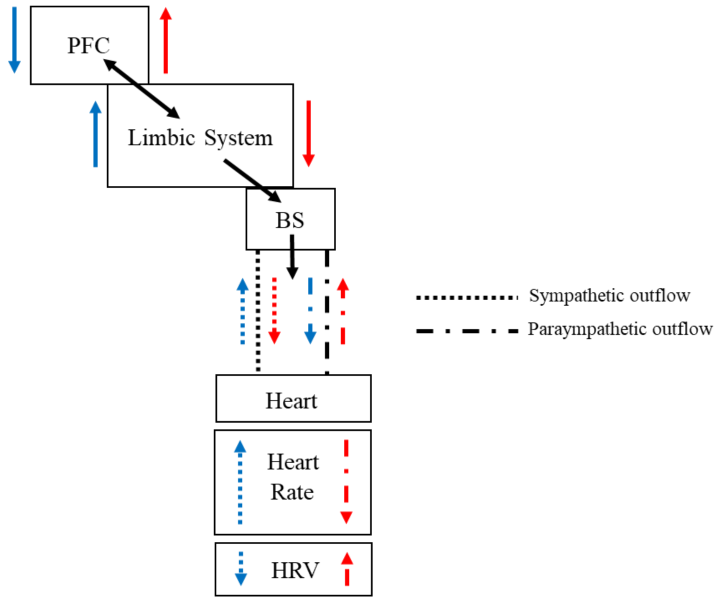
One would expect non-insight problem solving, typically defined as conscious and effortful, to be closely related to the trait- and resource-dependent perspectives of prefrontal-based WMC. Specifically, WMC should be positively associated with non-insight problem solving, and non-insight should depend on the availability of cognitive resources. However, for insight, the relation is less obvious. If insight entails a conscious solution search, it should also relate to the trait- and resource-dependent perspectives of prefrontal-based WMC, like non-insight. However, if insight depends more on unconscious processes, it should be relatively independent of both perspectives. The main aim of this study is to address to what extent insight and non-insight are (differentially) related to prefrontal-based WMC from both a trait- and resource-dependent perspective.

1.2 The neurovisceral integration model: Prefrontal functionality and vagally mediated HRV

One way to approach this conundrum is by assessing heart rate variability (HRV; Laborde et al., 2018). HRV refers to the variability in the time interval between two consecutive heartbeats (i.e., inter-beat intervals or IBI). HRV stems from the dynamic interplay between the parasympathetic (i.e., PNS; rest and digest) and sympathetic nervous system (i.e., SNS; fight or flight), where the former decreases heart rate (HR) and the latter increases it. More importantly, prefrontal functionality is often indexed using HRV (Laborde et al., 2018; Thayer et al., 2009). One of the theories linking prefrontal functioning to HRV is the neurovisceral integration model (Thayer et al., 2009). This model postulates that the prefrontal cortex modulates the heart's activity through distinct cortical-subcortical pathways. Namely, the prefrontal cortex regulates and tonically inhibits the activity of the limbic system, which has an inhibitory effect on the PNS outflow and an excitatory effect on SNS outflow (Smith et al., 2017a; Thayer et al., 2009). Thus, increased prefrontal resource availability is expected to be associated with increased inhibition of the limbic system. Consequently, this inhibition of the limbic system results in expanding PNS outflow and condensing SNS outflow, thereby causing HR to decrease and HRV to increase. The opposite pattern is expected for decreased prefrontal resource availability (see Figure 1 for an elementary schematic representation). The PNS is regarded as the HR's dominant controller, propagating its outflow via its vagus nerve (Laborde et al., 2018). PNS functions as a brake continuously engaged to inhibit the tonically active SNS outflow and the heart's intrinsic beating rate (i.e., 100-110 bpm). Engaging this vagal brake increases HRV by decreasing the HR to conserve and/or restore resources, which is the favoured bodily state (Laborde et al., 2018; Shaffer & Ginsberg, 2017). As such, the prefrontal cortex modulation of the heart's activity will mainly operate through this vagal brake. Therefore, HRV mediated via this vagal brake, instead of HR, is argued to be the essential index reflecting prefrontal resource availability and consumption (Laborde et al., 2018).

Figure 1

Schematic Representation of the Neurovisceral Integration Model



Note. PFC, prefrontal cortex; the limbic system, involving cingulate cortex, insula, amygdala, and hypothalamus; BS, brain stem; HRV, heart rate variability; blue arrows indicate the cascading actions if PFC resource availability decreases; red arrows indicate the cascading actions if PFC resource availability increases; black arrows show interactions between the different brain structures and their output to the heart. This schematic figure is an elementary representation of the Neurovisceral Integration Model (for more detailed information about this model, see Thayer et al., 2009).

1.3 Vagally mediated HRV: A trait and resource-dependent perspective

Hence, prefrontal cortex functionality and vagally mediated HRV (i.e., vmHRV) are closely related. This relationship has been approached from two perspectives (Laborde et al., 2018; Smith et al., 2017). Namely, as an index to size up people's prefrontal resource availability to address environmental challenges (e.g., to solve problems; Smith et al., 2017; Thayer et al., 2009) and as an index of prefrontal resource consumption (e.g., due to mental effort exertion; Laborde et al., 2018). Based on this, vmHRV is a suitable parameter to index prefrontal functionality from a trait- and resource-dependent perspective. From a *trait* perspective, vmHRV measured during a resting-state is considered to reflect people's trait vmHRV, a relatively stable feature across time (Bornstein & Suess, 2000; Li et al., 2009). For instance, it has been found that higher resting-state vmHRV is associated with enhanced executive functioning in general (Forte et al., 2019; Jennings et al., 2015; Magnon et al., 2022) and improved inhibitory control more specifically (Kimhy et al., 2013; Ottaviani et al., 2018). These studies illustrate that trait vmHRV is positively associated with cognitive outcomes. A recent study by Zeng et al. (2023) has further substantiated these findings by demonstrating that increased performance of individuals with high resting-state vmHRV on a WM task was related to higher efficiency in regulating neural resources in the prefrontal cortex, as compared to their lower vmHRV peers. From a *resource*-dependent perspective, vmHRV reactivity during task performance and vmHRV rebound during a post-task recovery period are indices used to represent prefrontal resource consumption. Research has shown the heart's reactivity during task performance (e.g., Cranford et al., 2014; Hansen et al., 2003) and during the application of non-invasive brain stimulation (Makovac et al., 2017; Nikolin et al., 2017). For instance, Nikolin et al. (2017) showed that excitatory brain stimulation to the left dorsolateral prefrontal cortex increased the power of the signal in the high-frequency band (i.e., a vmHRV parameter) compared to a sham condition. This showed that enhancing prefrontal resources affects vmHRV. Furthermore, HRV tends to recover to its resting-state value following an event-related HRV decrease (Balzarotti et al., 2017; Smith et al., 2020). The studies discussed above provide ample support for the premises put forward by the neurovisceral integration model regarding the bi-directional connection between the brain and the heart.

1.4 The heart's activity in insight/non-insight problem solving

Although this shows that vmHRV might be a valuable index to study how insight and non-insight problem solving are (in)differentially related to WMC from both a trait- and resource-dependent perspective, only a few studies have assessed heart function to determine how insight and non-insight problem solving are related to WMC (e.g., Jausovec & Bakracevic, 1995; Shen et al., 2017). Of these studies, to our knowledge, the study of Shen et al. (2017) is the only one that assessed vmHRV for both solution types. They presented participants with the compound remote associates test (i.e., CRA), where participants receive three cue words (e.g., fox/man/peep), and they have to search for a fourth compound solution word (e.g., hole) to form three new meaningful compound words (e.g., fox*hole*/man*hole*/peep*hole*). After each successfully solved CRA word puzzle, participants indicated, based on their subjective experience, whether they solved it with insight or non-insight. vmHRV, more specifically the root mean square of successive differences between normal IBIs (RMSSD), was analyzed in the last four seconds preceding solution retrieval. No difference was observed between resting-state baseline RMSSD and RMSSD preceding solution retrieval for both solution types. However, a negative association was found between the participants' average solution time and RMSSD measured before solution retrieval, solely for the CRA word puzzles solved with insight. They concluded that, for insight solutions, an increased average solution time is associated with increased recruitment of prefrontal resources to restructure the solution space immediately before finding the solution. This would imply that insight is relatively independent of WMC until the time window immediately preceding solution retrieval. Based on HR, Jausovec and Bakracevic (1995) also found a differential pattern in the heart's functioning for insight and non-insight. They presented participants with insight and non-insight problems and continuously tracked HR during the solution search. Their results showed that for non-insight problems, the trajectory of the HR displayed a gradual increase until solution retrieval. For the insight problems, the trajectory of the HR suddenly increased immediately before solution retrieval (see also Lackner et al., 2013). This observation implies that solving non-insight problems relied on WMC, increasingly depleting the availability of prefrontal resources. Contrarily, solving insight problems was relatively independent of WMC and, therefore, prefrontal resources were mainly unaffected. Overall, these studies suggest that vmHRV might be differentially affected for insight and non-insight due to a discrepancy in reliance on WMC.

Although these studies certainly paved the way, they were not able to fully address this research question. First, Jausovec and Bakracevic (1995) only used two insight and two non-insight problems, leading to a very limited amount of data points per participant. Furthermore, these problem types were also very different in nature (i.e., two visuospatial riddles and two simple verbal math problems, respectively), hampering a direct comparison between the two solution strategies. Furthermore, recent studies have shown that such insight problems can sometimes be solved with non-insight and such non-insight problems with insight (Danek et al., 2016; Webb et al., 2016). A proposed solution for this is to rely on participants' subjective self-reports to classify solved problems into those solved with insight and non-insight (see Bowden & Jung-Beeman, 2007; Shen et al., 2017). Second, although Shen et al. (2017) addressed this insight/non-insight problem type issue by using insight/non-insight self-reports, they only assessed vmHRV in the last four seconds of the solution search. The time to solve the CRA word puzzle varied in length for each trial (i.e., between 4s and 30s). Therefore, the last four seconds sometimes reflected the full solution search and sometimes only the end. We contend that this four-second vmHRV interval may not only reflect different aspects of the solution search depending on trial length but also leaves a large part of the solution search unexplored for some trials. This unexplored part might be crucial when determining insight and non-insights' reliance on WMC. Specifically, an essential part of the solution search that might tax WMC occurs between problem representation and the moment of solution retrieval. Therefore, we argue for a consistent approach for each CRA trial where the full solution search interval is taken into account. Third, Shen et al. (2017) only assessed vmHRV reactivity during task performance but not during a post-task recovery period. For non-insight, prefrontal resources are expected to be taxed during the solution search, and prefrontal resource recovery, as indexed by vmHRV, should be expected afterward. If insight is similar to non-insight, the same pattern should emerge. However, if insight is a more automated unconscious process, prefrontal resources should remain largely unaffected during the solution search, and subsequent prefrontal resources recovery should be minimal or absent. Lastly, both Jausovec and Bakracevic (1995) and Shen and colleagues (2017) tackled this research question solely from a resource-dependent perspective. We argue that approaching this research question from a trait perspective would shed valuable light on how inter-individual differences in prefrontal resource availability might affect insight/non-insight problem solving performance.

1.5 Aims and hypotheses of the current study

The current study aimed to assess how prefrontal functionality, strongly related to WMC, is differentially associated with insight and non-insight problem solving by measuring vmHRV, as indexed by RMSSD (from now on both terms are used interchangeably), from two perspectives. First, in contrast to the studies mentioned above (Jausovec & Bakracevic, 1995; Shen et al., 2017), we approached vmHRV from a trait perspective to elucidate how inter-individual differences in prefrontal functionality relate to both solution types. To our knowledge, this is the first study using this approach within this field. Second, extending the studies mentioned above, we approached vmHRV from a resource-dependent perspective to clarify whether and how prefrontal resource engagement during task performance and its subsequent recovery differ between both solution types. To achieve these aims, we used CRA word puzzles in line with Shen et al. (2017) and relied on participants' subjective self-reports to classify solved word puzzles into those solved with insight and with non-insight. RMSSD was measured before, during, and after the full solution search.

Regarding the *trait* perspective of vmHRV, we hypothesized that non-insight solution search performance would be positively relate to resting-state RMSSD. Following the neurovisceral integration model, individuals with higher resting-state RMSSD should display enhanced prefrontal functionality, reflecting a better regulation of cognitive control functions to consciously solve problems, which should be vital for non-insight. If insight solution retrieval similarly relies on a conscious process, it should also display an association with resting-state RMSSD. However, if insight entails a more unconscious process, it would not depend on the individual's intrinsic prefrontal functionality to the same extent as non-insight. In that case, insight solution search performance would be associated less or not at all with resting-state RMSSD.

Concerning the *resource*-dependent perspective of vmHRV, we predicted that non-insight should be associated with a decrease of RMSSD during the solution search relative to its resting-state measurement. In line with the neurovisceral integration model, consciously engaging cognitive control functions during the problem-solving process taxes prefrontal resources, which is expected to disengage the vagal break causing RMSSD to decrease. Furthermore, we expected that for non-insight, RMSSD would recover to baseline level following the solution-search-related RMSSD decrease during a recovery period. Contrarily, if insight results from more unconscious processes largely independent of the engagement of prefrontal-based cognitive control functions, RMSSD during solution search and recovery is expected to remain relatively stable from baseline to solution search to recovery.

2. METHOD

2.1 Participants

A convenience sample of 84 undergraduates took part in this study. They received course credits for their participation. Several inclusion criteria were used: non-smokers, body-mass index <30, Beck Depression Inventory score <29, no cardiovascular or neurological medication use, and no history of or current cardiopulmonary diseases, psychiatric disorders and/or neurological disorders. Moreover, participants had to adhere to several instructions concerning their daily routines immediately preceding their participation: no alcohol consumption the night before and on the day of the experiment, at least six hours of sleep the night before the experiment, and no caffeine consumption, no heavy meal consumption and no strenuous physical activity two hours before the experiment. We used these inclusion criteria and daily instructions as it has been indicated that they might negatively influence the ECG recording (see Quintana & Heathers, 2014).

Of the 84 participants, seven were excluded due to technical issues with the ECG recording (i.e., incorrect electrode placement, $N = 1$, disconnection of the electrodes during recording, $N = 2$, and trigger interface unresponsiveness, $N = 4$). Five participants were also excluded after visual inspection of the ECG signal because of an abnormal ECG signal (see Appendix A for an example; Kumral et al., 2019; Shaffer & Ginsberg, 2017). For the remaining 71 participants, the percentage of to-be-excluded ECG noise epochs (i.e., an epoch with an uninterpretable ECG wave morphology) and the percentage of corrected IBIs in the ECG noise-free part were determined (see data preprocessing for a detailed explanation). If one or both of these percentages surpassed the 5% poor-data-quality threshold in the baseline ECG recording, the participant was excluded from the analysis (see Munoz et al., 2015 for a similar procedure). Based on these criteria, we excluded another three participants. Lastly, we omitted participants with severely outlying HRV data in the baseline ECG recording. This criterion was determined based on Tukey's (1977) method of three interquartile ranges above the sample median (i.e., RMSSD, $Mdn = 33.30$ms; $IQR = 18.43$ms; range 0 – 88.59ms; see Kumral et al., 2019 for a similar procedure). This led to the additional exclusion of one participant. The final sample consisted of 68 participants (mean age = 19, $SD = 1.29$, 62 female, mean BDI score = 8.53, $SD = 5.73$), which is more than three times the size of previous similar insight studies (Jausovec & Bakracevic, 1995, $N = 19$; Shen et al., 2017, $N = 22$). All participants provided their written informed consent before they participated. As HRV studies on insight are scarce, we used a medium effect size to estimate the power, which is generally considered a valid effect size in psychological research (Brysbaert & Stevens, 2018). The current sample size is sufficient to detect medium effect sizes ($\eta^2 = 0.07$) with a statistical power of 80% (Campbell & Thompson, 2012). The Social and Societal Ethics Committee (SMEC) of the KU Leuven approved the study (approval number G-2019 12 1929).

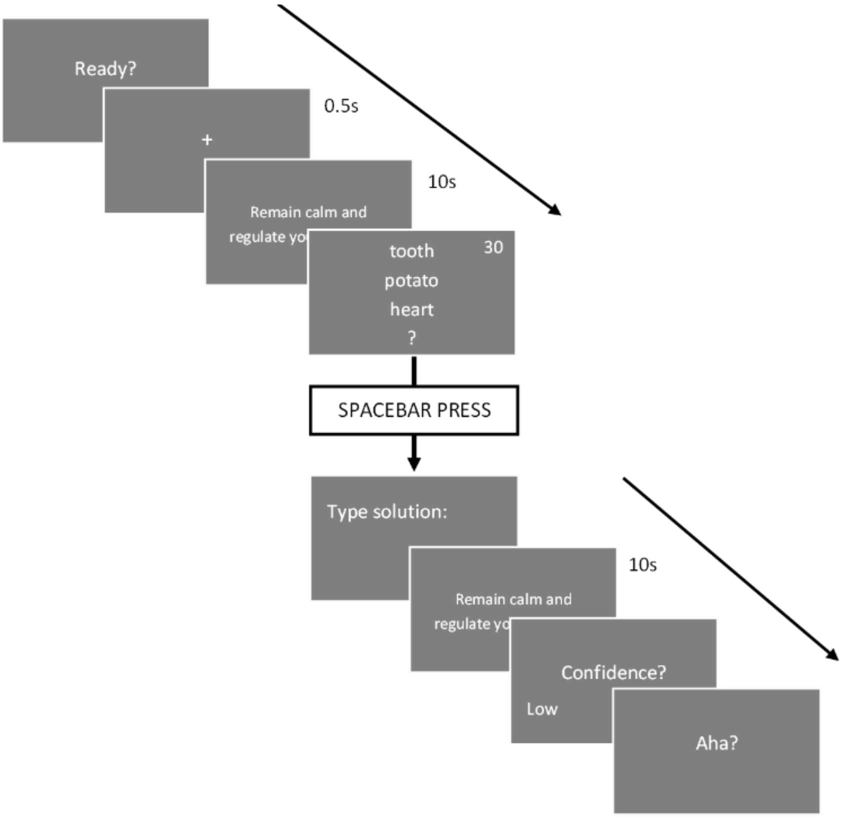
2.2 Assessment and Measurement

To measure insight and non-insight using the same type of problems, the Dutch version of the Compound Remote Associates Test (CRA) was used (Stuyck et al., 2021; https://osf.io/snb3k/). In the CRA, participants are presented with a word puzzle on each trial containing three words. They are requested to find a fourth word to attach to each of these three words to make three new compound words (e.g., *escape* hatch, escape route, *escape* artist). The Dutch version of the CRA contains 76 word puzzles (i.e., 70 experimental and 6 practice CRA trials). For the experimental trials, half of the word puzzles required a solution word to be attached to the back of each of the three words of the puzzle, and the other half required a solution word to be attached to the front.

Figure 2 depicts an example of a CRA word puzzle trial. The experiment was self-paced. After participants indicated their readiness with a spacebar press, a fixation cross was presented at the screen's center for 0.5s. After that, there was a 10s rest interval during which participants were instructed to remain calm and regulate their emotions (see Shen et al., 2017 for a similar procedure). Next, the CRA word puzzle was presented for maximal 30s. All words of the word puzzle were presented in white on a silver background (i.e., typeface Arial and letter height 1.5/10 of the screen size). The three words of the word puzzle were presented vertically on top of each other with a question mark beneath it (i.e., screen's y-axis ranging from -1 bottom to 1 top; first word on 0.45, second word 0.25, third word on 0.05, and the question mark on -0.25) at the center of the screen's x-axis. A countdown timer was presented in the screen's upper right corner. After indicating that a solution was found with a spacebar press, participants were presented with a screen for a maximal duration of 10s to type in their solution. After that, there was a trial-by-trial HRV recovery interval of 10s, during which participants were instructed to remain calm and regulate their emotions. Next, a screen was presented, probing participants which solution type had led to the solution. They answered this question by pressing (1) for insight (Aha!), (2) for non-insight (without Aha!), or (3) for another strategy. There was no time limit to respond to this solution-type question. After that, participants were asked to report their solution confidence by moving a scale slider with their mouse to select a position on a horizontal scale (i.e., vertical white rectangular bar) between low (0) and high (1) confidence. The starting position of the slider on the horizontal scale randomly varied on each trial. There was no time limit to respond to this solution-confidence question. After each solved CRA trial, the solution-type and solution-confidence questions were presented in random order. If participants could not solve the CRA word puzzle in the allotted time, the next CRA trial immediately began after the time had run out.

Figure 2

Example of a CRA Word Puzzle Trial



2.3 Equipment

Participants were seated individually in a quiet, dimly lit room held at a constant temperature between 21° and 23°. They faced the computer monitor from approximately 60cm. A Dell Optiplex 3060 computer was used with a Dell 23.6-inch monitor. The experiment was programmed with PsychoPy v2021.2.3 (Pierce et al., 2019). Nexus-10 MKII (Mind Media BV, Herten, the Netherlands) was used as the ECG recording device (CE-certified; 93/42/EC Annex XII). The device obtains the ECG signal in microvolts with a sampling rate of 256 Hz. Three pre-gelled Ag/AgCl electrodes were used. Following the modified Lead-II placement, these were attached to the upper body (see Kuipers et al., 2017 for a similar procedure). Namely, the negative electrode below the center of the right collarbone, the positive electrode on the lower left rib cage and the ground electrode below the left collarbone. Before placing the electrodes, the skin was cleaned with an alcohol pad. The data obtained by the Nexus-10 MKII (i.e., ECG signal) were preprocessed using Kubios premium (v. 3.4.2; Tarvainen et al., 2014).

2.4 Procedure

All participants were examined between 9 am and 5 pm. Before entering the laboratory, participants were asked to go to the toilet if needed. The experiment leader explained that a full bladder and/or bowel might influence the HRV recording. After that, participants' eligibility to participate in the experiment and their adherence to the instructed daily routines were checked. The BDI was only administered after the experiment to avoid any influence on the participants' HRV recording due to the valence of the questions. Subsequently, participants received instructions on how to attach the electrodes to the upper body. This instruction was accompanied by an image displaying the modified Lead-II placement. After participants had attached the electrodes, the experiment leader visually checked if the electrodes were attached correctly. Participants also wore a wrist-worn device (i.e., Empatica E4) on their non-dominant hand in light of another validation study (Stuyck et al., 2022). Participants were instructed to remain in an upright seated position without crossing their legs and refrain from excessive movement (e.g., coughing, arm-stretching). It was explained that any deviation from these instructions might negatively influence the ECG recording. Next, participants performed that CRA test as described above. Before initiating the CRA test, elaborate instructions were provided so that participants fully understood the aim of a CRA word puzzle, what each solution type entails, and how to respond to the different questions. To explain the insight and non-insight concepts, we relied on definitions employed in previous research (Danek et al., 2014; Hedne et al., 2016; Jung-Beeman et al., 2004). The detailed instructions can be found in Appendix B. During the practice trials, feedback was provided in the form of the correct solution.

To assess the association between trait vmHRV and behavioural performance on both solution types, a baseline HRV recording was obtained after the six practice CRA trials. Specifically, a 10 min resting interval followed these practice trials. During this interval, participants were instructed to relax and control their emotions to minimize the likelihood of engaging in ruminative or emotionally valenced thoughts that might affect ECG recording. To minimize changes in their behaviour related to the awareness of the recording, during this 10 min interval, we told participants that this interval was meant to return to a relaxed state and that we were mainly interested in their (later) CRA word puzzle performance. This 10min baseline interval consisted of a 5 min acclimatization period and a subsequent 5 min baseline interval. This 5 min baseline interval was used to measure the trait vmHRV and functioned as the baseline measurement to assess changes in vmHRV related to the solution search for each solution type. After the 10 min resting interval, participants received a summary of the instructions followed by the 70 experimental CRA trials.

2.5 Vagally mediated HRV: RMSSD

To assess vmHRV, we used the root mean square of successive differences between normal IBIs (i.e., RMSSD), a time-domain HRV metric reflecting the variability in beat-to-beat intervals (Shaffer & Ginsberg, 2017). RMSSD is considered a suitable candidate for reflecting PNS's effect on HRV (Laborde et al., 2017). High-frequency band power is another HRV metric linked to the PNS effect on HRV, but it is more affected by respiration compared to RMSSD (Laborde et al., 2017; Shaffer & Ginsberg, 2017). Given the high correlation between high-frequency band power and RMSSD (i.e., $r = .86$ in the current sample), we only used RMSSD to represent vmHRV.

2.6 Data preprocessing

The Kubios HRV software was used to preprocess the ECG data (Tarvainen et al., 2020). We used the automatic artifact correction algorithm of Kubios to correct for potential artifacts in the IBI time series (e.g., ectopic beats, missed and extra beats). All detected artifacts were subsequently replaced with IBIs based on cubic spline interpolation. Additionally, we applied a visual inspection of the ECG signal to identify unstable recording epochs, missed r-peaks, and missed artifacts by an automatic correction algorithm. In case an artifact was detected, a manual correction was applied. We only retained data that was minimally 95% noise free and had a maximum of 5% corrected IBIs. In Appendix C, a detailed description of the preprocessing steps can be found.

To create the solution search and recovery intervals over which RMSSD was calculated, we merged the time intervals of the individually correctly solved CRA word puzzle trials for each solution type. This led to four RMSSD observations per participant (i.e., solution search insight, solution search non-insight, recovery insight and recovery non-insight). These merged solution-search intervals consisted of a series of time intervals with differing lengths, as CRA word puzzles took between 1s and 30s to be solved. The merged recovery intervals always consisted of a series of time intervals of 10s, as the recovery time interval after each trial had a fixed length of 10s. The minimum solution search and recovery interval length deemed acceptable for assessing RMSSD was 10s. Previous research (e.g., Munoz et al., 2015) has shown that RMSSD calculated with a 10s ECG recording gives a reliable approximation of the RMSSD obtained with a 5min ECG recording.

2.7 Statistical Analysis

For all statistical analyses, we only included the CRA word puzzles solved with insight and non-insight. The solved CRA word puzzles where participants reported having used "another strategy" were not included because they have no informational value regarding the solution types of interest. This led to the omission of 252 of the 3355 CRA word puzzle trials.

2.7.1 Trait vagally mediated HRV

To assess the association between the trait vmHRV and CRA performance (i.e., solution time, solution accuracy, solution confidence, and the number of correctly solved CRA word puzzles) for each solution type, we relied on the RMSSD obtained with the 5 min baseline ECG recording. Next, we used three (generalized) linear mixed models ([G]LMM) and one generalized linear model (GLM) to assess our hypotheses. These GLMMs and GLM contained solution type (i.e., insight vs. non-insight), RMSSD (i.e., continuous predictor), and their interaction term as fixed effects. The first LMM had solution time as a continuous outcome variable and was solely based on the correctly solved CRA word puzzles (Baayen & Milin, 2010). We applied a Box-Cox transformation to solution time ($\lambda = -0.2$) to accommodate non-normality and heteroscedasticity (i.e., $\text{BoxCox solution time} = \frac{\text{solution time}^{-0.2}-1}{-0.2}$; Atkinson et al., 2021). Afterward, solution time was back-transformed for interpretability (Atkinson et al., 2021). A second GLMM was used with solution accuracy as a binary outcome variable (i.e., 0 =

incorrect and 1 = correct; Sommet & Morselli, 2017). The third GLMM contained solution confidence as a bounded outcome variable and was solely based on the correctly solved CRA word puzzles (i.e., .005 - .995; Verkuilen & Smithson, 2012). A fourth and final GLM was used with the number of correctly solved CRA word puzzles as a count outcome variable (Gardner et al., 1995). In this GLM, each participant had one count observation for insight and one for non-insight. To account for the non-independence in the GLM (i.e., two observations clustered in participants), we used robust standard errors (Zeileis et al., 2020). In all specified models above, RMSSD was standardized across participants by rescaling it to *z*-scores to enhance interpretability and convergence of the (G)LMMs and GLM (Enders & Tofighi, 2007). Furthermore, we used sum coding to set the contrasts, coding insight as 0.5 and non-insight as -0.5 (Schad et al., 2020). The Satterthwaite approximation method was used to assess the significance of the LMM, while the Wald test assessed the significance of the GLMMs and the GLM. All (G)LMMs included a random intercept for the participant and the CRA word puzzle, taking into account between participant and between CRA word puzzle variations (Baayen et al., 2008).

If a significant effect was found for solution type, we conducted a post-hoc test to extract the model's estimates for insight versus non-insight, accompanied by Cohen's *d* effect size. If there was a significant effect of scaled baseline RMSSD, the beta coefficient of the model represented the effect size and was used for interpretation. For the LMMs of solution time and solution confidence, this is straightforward. However, this is less clear for the GLMM of solution accuracy and the GLM of the number of correctly solved CRA word puzzles. To address this, we exponentiated the beta coefficients to obtain *odds ratios* (*OR*) for the GLMM of solution accuracy and *incidence rate ratios* (*IRR*) for the GLM of the number of correctly solved CRA word puzzles. An *OR* above or below one indicates an increase or decrease in the likelihood of solving a CRA word puzzle correctly, respectively (Sommet & Morselli, 2017). We divided one by the *OR* in case the *OR* was below one to express them as "times less likely". An *IRR* reflects the multiplicative factor with which the outcome variable increases or decreases with a one-unit increase of the predictor variable (Wilson, 2022). If there was a significant interaction effect between solution type and scaled baseline RMSSD, a post-hoc test was used to estimate the slope coefficients and their significance for both solution types.

2.7.2 The resource-dependent vagally mediated HRV

We used different interval types to assess vmHRV's reactivity to the solution search and its reactivity during a post-trial recovery period. Specifically, we used the same 5 min ECG recording of the trait vmHRV analysis to calculate *baseline* RMSSD (see Shen et al., 2017 for a similar procedure). Next, only the correctly solved CRA word puzzle trials were considered for the merged *solution-search* and *recovery* intervals (N = 2420). Of these correctly solved CRA word puzzles, 51 trials could not be used due to a technical issue with the trigger interface, and 39 trials were omitted because of an uninterpretable ECG signal morphology (i.e., noise epochs). So the calculation of RMSSD for the merged *solution search* and *recovery* intervals was based on 2330 correctly solved CRA word puzzle trials.

To assess the reactivity of vmHRV in the baseline, solution search, and recovery intervals, we used an LMM with solution type (i.e., two levels: insight and non-insight), interval type (i.e., three levels: baseline, solution search and recovery), and their interaction term as fixed effects. In this LMM, RMSSD was the continuous outcome variable (Baayen & Milin, 2010). To accommodate non-normality and heteroscedasticity, we used a Box-Cox transformation for RMSSD with λ = 0.3 (i.e., $\mathrm{BoxCox\ RMSSD} = \frac{\mathrm{RMSSD}^{0.3}-1}{0.3}$; Atkinson et al., 2021). Afterward, we back-transformed RMSSD for the interpretability of the results (Atkinson et al., 2021). Except for the specified characteristics above, this LMM was similar to the LMM of solution time of the trait vmHRV. We used model comparison based on the likelihood ratio test (χ^2) to extract the main and interaction effects of the LMM. Namely, we compared the full model to three reduced models that either excluded one of the main effects or the interaction effect. Finding statistical significance in those comparisons implies the presence of main and/or interaction effects (Levy, 2018).

In case of a significant main and/or interaction effects, we applied a post-hoc test to interpret them further. For the main effect of solution type, we contrasted insight versus non-insight. If there was a main effect of interval type, we conducted a pairwise comparison for the three interval types (i.e., baseline vs. solution search, baseline vs. recovery, solution search vs. recovery). In case of a significant interaction effect, we compared the three interval types pairwise conditionally on the solution types. The Tukey method was used to adjust for multiple comparisons. Cohen's *d* effect sizes accompanied all pairwise contrasts.

The open-source R language to perform statistical analysis was used (R Core Team, 2020). Details of the package used for the statistical analysis can be found in Appendix D. All data and program code (R and PsychoPy) are placed on the OSF platform (https://osf.io/7hp5u/).

3. RESULTS

3.1 Trait vagally mediated HRV

Together, participants solved 3103 CRA word puzzles (i.e., insight = 1735 and non-insight = 1368), of which they correctly solved 2420 CRA word puzzles (i.e., insight = 1541 and non-insight = 879). Based on these correctly solved CRA word puzzles, participants, on average, solved 23 CRA word puzzles with insight and 13 with non-insight. The baseline measurement of RMSSD showed, across participants, an average RMSSD of 36ms (SD = 13, range 15 – 68). The descriptive statistics of the outcome variables can be found in Table 1.

Table 1

Descriptive Statistics of the Outcome Variables

	Insight		Non-Insight	
	M(*SD*)	range	*M*(*SD*)	range
Solution time	8.36(5.86)	1.58-29.68	10.67(6.67)	1.26-29.46
Solution accuracy	89%(10%)	60%-100%	64%(23%)	0%-100%
Solution confidence	.83(.19)	.005-.995	.67(.26)	.005-.995
#Solved	23(9)	1-43	13(7)	1-38

Note. Solution time is expressed in seconds; Solution accuracy is the percentage of correctly solved CRA word puzzles; Solution confidence is rescaled to range from .005 low confidence to .995 high confidence; #Solved, is the number of correctly solved CRA word puzzles.

3.1.1 Solution time

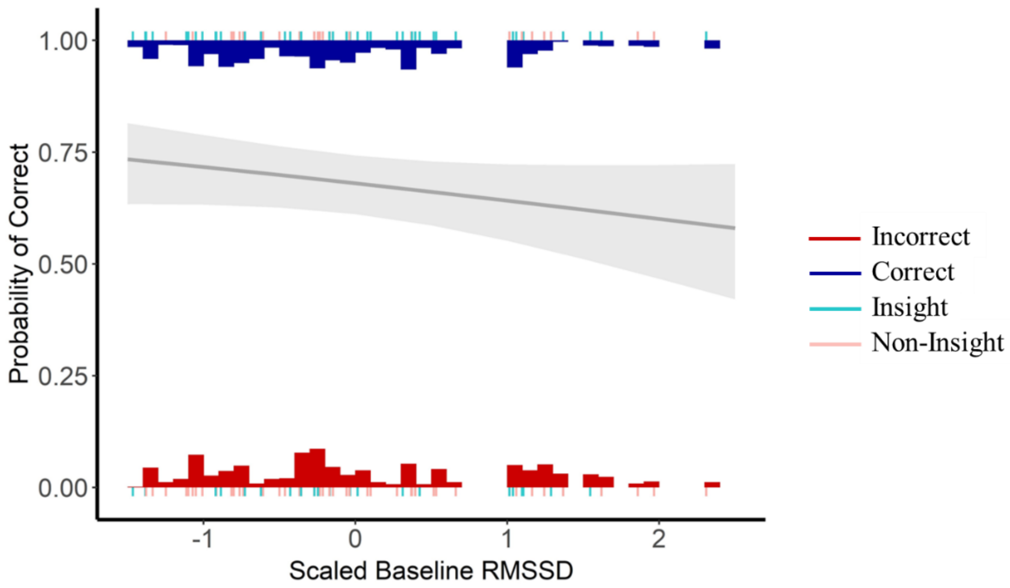
Based on the correct CRA trials, an LMM was constructed with (Box-Cox transformed) solution time as the outcome variable and solution type (i.e., insight vs. non-insight), scaled baseline RMSSD (continuous variable), and their interaction term as fixed effects (Table 1 in Appendix E depicts the model). Solution type was a significant predictor of solution time, $t(2390) = -9.31$, $p < .001$. CRA word puzzles solved with insight were solved significantly faster (M = 6.98s) than those solved with non-insight (M = 8.89s), Cohen’s d = 0.43 (95% CI [0.34, 0.52]). The scaled baseline RMSSD and the interaction between solution type and the scaled baseline RMSSD were not significant, p = .999 and p = .256, respectively.

3.1.2 Solution accuracy

Based on the correct and incorrect CRA trials, a similar GLMM was built with solution accuracy as a binary outcome variable (Table 2 in Appendix E depicts the model). Solution type was a significant predictor of solution accuracy, $Z = 15.07$, $p < .001$. CRA word puzzles solved with insight ($M = 92\%$) were more likely to be correct than those solved with non-insight ($M = 68\%$), Cohen's $d = -1.74$ (95% CI [-1.97, -1.51]. The scaled baseline RMSSD was a marginally significant predictor of solution accuracy, $OR = 0.81$ [1/0.81 = 1.24], $Z = -1.85$, $p = .065$. This showed that a one-unit increase of the scaled baseline RMSSD made it 1.24 times less likely to solve CRA word puzzles correctly. Figure 3 illustrates this result. The interaction between solution type and the scaled baseline RMSSD was not significant, $p = .436$.

Figure 3

Scaled Baseline RMSSD Predicting Solution Accuracy



Note. X-axis, Scaled Baseline RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; Y-axis, predicted probability of solving a CRA word puzzle correctly; bars at the bottom of the graph represent the percentage of incorrectly solved CRA word puzzles for a given bin of scaled RMSSD; bars at the top of the graph represent the percentage of correctly solved CRA word puzzles for a given bin of scaled RMSSD. To extract the percentage of correctly solved word puzzles, one should subtract the bar height from one (e.g., 1- bar height 0.9 = 0.1 or 10%); rugs at the bottom of the graph represent CRA word puzzles incorrectly solved with insight and non-insight at the observational level; rugs at the top of the graph represent the CRA puzzles correctly solved with insight and non-insight at the observational level; grey line, association between probability of correctly solved CRA word puzzles and scaled RMSSD; grey shade, represents the 95% confidence interval.

3.1.3 Solution confidence

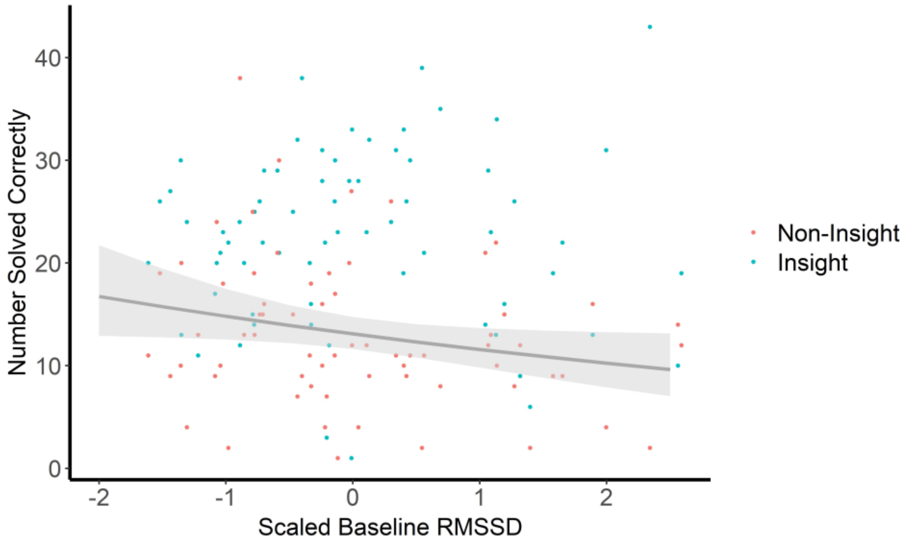
Similarly, based on the correct CRA trials, a GLMM was built with solution confidence as a bounded outcome variable (range = .005 - .995; Table 3 in Appendix E illustrates the model). Solution type was a significant predictor of solution confidence, $Z = 14.42$, $p < .001$. Solution confidence was significantly higher for word puzzles solved with insight ($M = .81$) than those solved with non-insight ($M = .69$), Cohen's $d = -0.13$ (95% CI [-0.15, -0.12]. The scaled baseline RMSSD and the interaction between solution type and the scaled baseline RMSSD were not significant, $p = .474$ and $p = .496$, respectively.

3.1.4 The number of correctly solved word puzzles

Lastly, a GLM was constructed with the number of correctly solved CRA trials as a count outcome variable (Table 4 in Appendix E shows the model). Solution type was a significant predictor of the number of correctly solved CRA word puzzles, $Z = 5.97$, $p < .001$. Correctly solved CRA word puzzles with insight ($M = 23$) were more frequent than those solved with non-insight ($M = 13$), Cohen's $d = -0.52$ (95% CI [-0.68, -0.36]). The scaled baseline RMSSD was also a significant predictor of the number of correctly solved CRA word puzzles, *incidence rate ratio* = 0.93 [1 – 0.94 = 0.06], $Z = -2.26$, $p = .024$. This showed that a one-unit increase of the scaled baseline RMSSD corresponded to a 6% decrease in the number of correctly solved CRA word puzzles. Figure 4 illustrates this result. The interaction between solution type and the scaled baseline RMSSD was not significant, $p = .181$.

Figure 4

Scaled Baseline RMSSD Predicting the Number of Correctly Solved CRA Word Puzzles



Note. X-axis, Scaled Baseline RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to *z*-scores; Y-axis, the number of correctly solved CRA word puzzles; grey line, represents the association between the number of correctly solved CRA puzzles and the scaled RMSSD; grey shade, represents the 95% confidence interval.

3.2 Resource-dependent vagally mediated HRV

The descriptive statistics of RMSSD and the length of the interval types used to calculate it are presented in Table 2 as a function of solution type (i.e., insight and non-insight) and interval type (i.e., baseline, solution search and recovery). This analysis is based on 2330 correctly solved CRA word puzzles.

Table 2

Descriptive statistics of the outcome variable RMSSD and the Length of Assessment Interval

		RMSSD		length of recording	
		M(*SD*)	range	*M*(*SD*)	range
Insight	baseline	35.32(13.08)	14.82-67.68	299s(2s)	286s-300s
	solution search	42.87(15.04)	17.00-73.68	186s(74s)	20s-389s
	recovery	42.76(15.36)	14.64-72.34	222s(84s)	10s-430s
Non-Insight	baseline	35.32(13.08)	14.82-67.68	299s(2s)	286s-300s
	solution search	42.29(16.57)	15.58-97.80	139s(74s)	11s-466s
	recovery	45.28(19.43)	12.11-89.78	130s(71s)	10s-387s

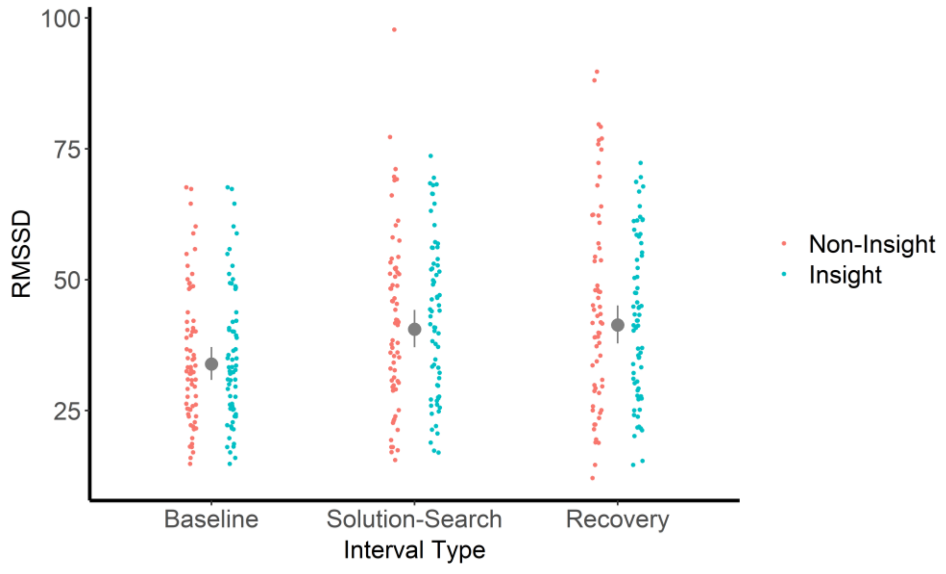
Note. RMSSD, root mean square of successive differences between normal inter-beat intervals expressed in ms; length of recording refers to the length of the interval type used to calculate RMSSD; the solution search and recovery intervals used to calculate RMSSD were constructed by merging all trial-by-trial intervals of the correctly solved CRA word puzzles.

3.2.1 Resource-dependent vagally mediated HRV

An LMM was constructed with (Box-Cox transformed) RMSSD as the outcome variable and solution type (i.e., insight vs. non-insight), interval type (i.e., baseline vs. solution search vs. recovery), and their interaction term as fixed effects (Table 1 in Appendix F depicts the model). There was only a significant effect of interval type, $\chi^2(2) = 85.26$, $p < .001$. Post-hoc tests showed that baseline RMSSD ($M = 33.88$ms) was significantly lower than the solution-search RMSSD ($M = 40.56$ms), $t(335) = 7.99$, $p < .001$, Cohen' s $d = -0.99$ (95% CI [-1.25, -0.73]) and the recovery RMSSD ($M = 41.34$ms), $t(335) = 8.87$, $p < .001$, Cohen' s $d = -1.10$ (95% CI [-1.36, -0.84]). The solution-search RMSSD was not significantly different from the recovery RMSSD, $p = .998$. This finding is also illustrated in Figure 5. The effect of solution type and the interaction effect between solution type and interval type were not significant, $p = .822$ and $p = .564$, respectively.

Figure 5

RMSSD Predicted by Solution Type and Interval Type



Note. RMSSD, root mean square of successive differences between normal inter-beat intervals; RMSSD was back-transformed from a Box-Cox transformation with $\lambda = 0.3$; bars represent the 95% confidence intervals.

3.3 Exploratory analyses

3.3.1 Differentiating insight from non-insight

We performed two additional exploratory analyses to explore the absence of a difference between insight and non-insight in the main analysis.

3.3.1.1 Exploratory analysis 1: Early and late solution retrieval. In this exploratory analysis, we took the length of the solution search into account (range 1.26s to 29.68s) to unveil differences between insight and non-insight in the trait- and resource-dependent analyses. CRA trials of differing lengths might be less comparable because the timing of the solution search phases (i.e., problem representation, search and solution retrieval) differs. As such, analysing all CRA trials together might have obscured differences between insight and non-insight. Therefore, we have performed a median split on the overall solution times. Based on the median of solution time (Mdn = 7.006s), we divided solved CRA word puzzles into an *early* (i.e., < Mdn = 7.006s) and a *late* (i.e., > Mdn = 7.006s) sample. The early sample had a median solution time of 4.5s (SD = 1.3s, range 1.3s – 7.002s), and the late sample had a median solution time of 12.3s (SD = 5.8s, range 7.009s – 29.7s). Subsequently, we reran all statistical analyses described above on the early and late samples separately.

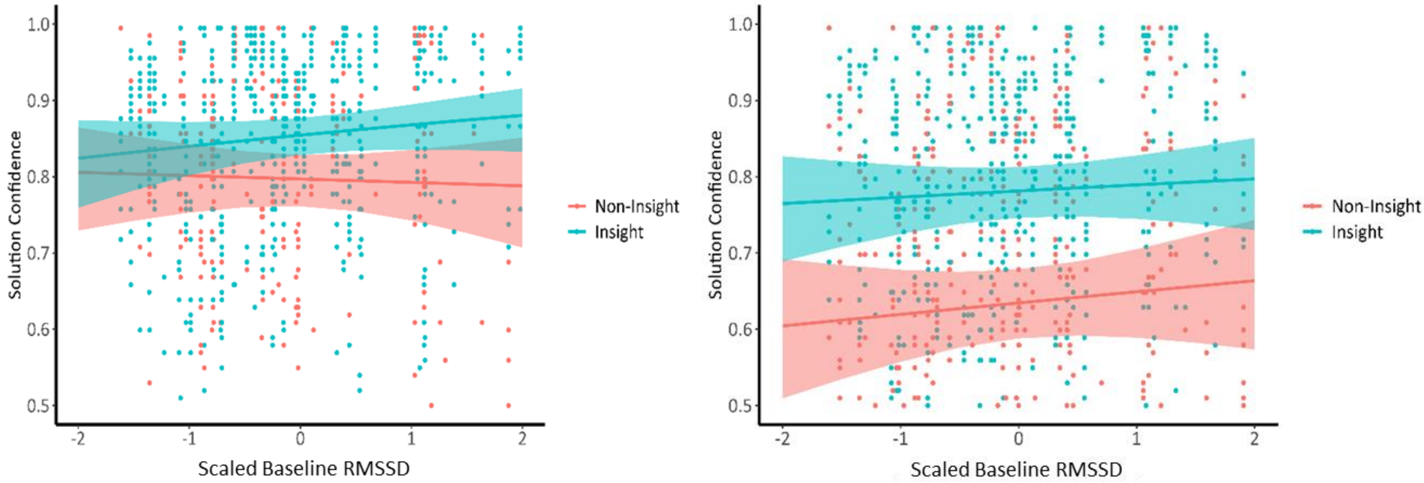
Trait vagally mediated HRV. For *solution time*, results were similar to the main analysis in both the early and late samples (see Table 1 in Appendix G for the models). Namely, only solution type was a significant predictor of solution time, $t(1145) = -3.65$, $p < .001$, and $t(1135) = -3.54$, $p < .001$, for the early and late samples, respectively. This result showed that word puzzles solved with insight were solved faster than those solved with non-insight for both the early and late samples (i.e., early: M = 4.61s versus 4.91s, Cohen's d = 0.26, 95% CI [0.12, 0.41]; late: M = 11.87s versus 12.66s, Cohen's d = 0.21, 95% CI [0.09, 0.33]). The scaled baseline RMSSD and the interaction between the scaled baseline RMSSD and solution type were not significant in both the early (i.e., p = .543 and p = .136, respectively) and late sample (i.e., p = .266 and p = .954, respectively). Note that the effect sizes for the main effect of solution type decreased from medium in the main analysis (i.e., Cohen's d = 0.43) to small (i.e., Cohen's d = 0.26 and 0.21 for early and late samples, respectively) in the exploratory analysis. This shows that splitting solution time into early and late samples made word puzzles solved with insight and non-insight more alike in terms of the length of the solution search.

For solution accuracy, solution type was a significant predictor for both the early (Z = 5.26, p < .001) and late (Z = 10.88, p < .001) samples, in line with the main analysis (see Table 2 in Appendix G for the models). CRA word puzzles solved with insight were more likely to be correct than those solved with non-insight for both the early and late samples (i.e., early: M = 96% versus 89%, Cohen's d = -1.17, 95% CI [-1.60, -0.73]; late: M = 86% versus 57%, Cohen's d = -1.58, 95% CI [-1.87, -1.29]). However, whereas scaled baseline RMSSD was a marginally significant predictor of solution accuracy in the main analysis, Z = -1.85, p = .065, this effect was now only observed for the late sample, OR = 0.79 [1/0.79 = 1.27], Z = -2.101, p = .036. This showed that a one-unit increase of the scaled baseline RMSSD made it 1.27 times less likely to solve CRA word puzzles correctly. Contrarily, on the early sample, the scaled baseline RMSSD was not a significant predictor of solution accuracy, p = .191. Lastly, similar to the main analysis, in both the early and late samples, the interaction effect between solution type and the scaled baseline RMSSD was not significant, p = .156 and p = .149, respectively.

For solution confidence, solution type was a significant predictor for both the early (Z = 6.20, p < .001) and late (Z = 11.80, p < .001) samples, in line with the main analysis (see Table 3 in Appendix G for the models). CRA word puzzles solved with insight received a higher solution confidence than those solved with non-insight for both the early and late samples (i.e., early: M = .86 versus .80, Cohen's d = -0.06, 95% CI [-0.08, -0.04]; late: M = .78 versus .64, Cohen's d = -0.17, 95% CI [-0.20, -0.14]). Similar to the main analysis, the main effect of the scaled baseline RMSSD was not significant in both the early and late samples, p = .619 and p = .475, respectively. However, unlike the main analysis, we did observe a significant interaction effect between solution type and the scaled baseline RMSSD, but only for the early sample, Z = 2.19, p = .028. Figure 7 illustrates this result, which shows that the difference in solution confidence between puzzles solved with insight and non-insight became more pronounced with increasing scaled baseline RMSSD. For the late sample, this interaction effect was not significant, p = .783.

Figure 6

Scaled Baseline RMSSD Predicting Solution Confidence for the Early and Late Samples



Note. left figure; early sample; right figure; late sample; X-axis, Scaled Baseline RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to *z*-scores; Y-axis, solution confidence (range .005 - .995); coloured lines, represent the association between solution confidence and the scaled RMSSD depending on solution type; coloured shade, represents the 95% confidence interval.

In order to conduct this exploratory analysis for *the number of correctly solved word puzzles*, we would need to analyse the data with generalized linear models due to only two observations being nested within participants. As generalized linear models, in contrast to generalized linear mixed models, do not handle missing values efficiently, participants with missing observations would be excluded from this, leading to biased estimates (Matuschek et al., 2017). Specifically, this would lead to the exclusion of nine participants in the early sample and two in the late sample. Furthermore, by calculating the number of correctly solved insight and non-insight word puzzles separately for the early and late samples, we would artificially split the total number of correctly solved word puzzles in two. This division is inconsistent with the idea of the total number of correctly solved word puzzles. Indeed, for some participants, there is an equal distribution of the number of correctly solved word puzzles in the early and late samples, but for other participants, this distribution is skewed, with either more word puzzles correctly solved in the early sample than in the late sample or vice versa. Therefore, we refrain from conducting the exploratory analysis for the number of correctly solved word puzzles.

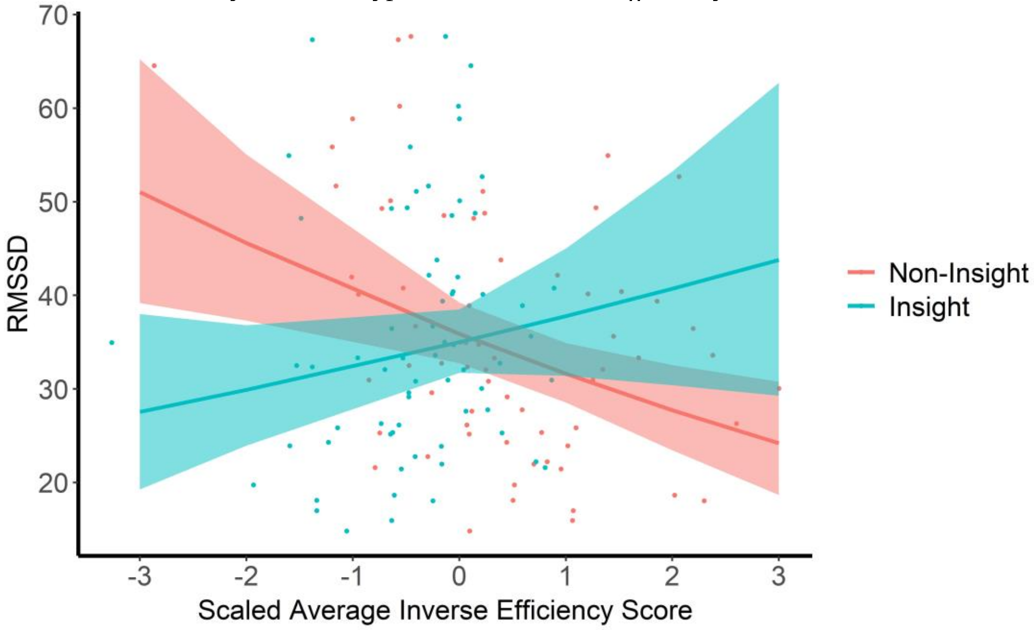
Resource-dependent vagally mediated HRV. For the resource-dependent approach to vagally mediated HRV, we observed similar results in the early and late samples as in the main analysis (see Table 4 in Appendix G for the models). There was only a significant effect of interval type in both the early and late samples, $\chi^2(2) = 51.83$, $p < .001$, and $\chi^2(2) = 51.47$, $p < .001$. Post-hoc tests showed that baseline RMSSD was significantly lower than the solution-search RMSSD for both the early and late samples (i.e., early: M = 33.63ms versus 39.50ms, $t(303) = 5.91$, $p < .001$, Cohen's $d = -0.79$, 95% CI [-1.06, -0.51]; late: M = 33.63 versus 39.31, $t(329) = 6.10$, $p < .001$, Cohen's $d = -0.76$ (95% CI [-1.01, -0.51]). The baseline RMSSD was also significantly lower than the recovery RMSSD for both the early and late samples (i.e., early: M = 33.63ms versus 40.07ms, $t(303) = 6.75$, $p < .001$, Cohen's $d = -0.87$, 95% CI [-1.12, -0.60]; late, M = 33.63ms versus 39.88ms, $t(330) = 6.65$, $p < .001$, Cohen's $d = -0.84$, 95% CI [-1.09, -0.58]). The solution-search RMSSD was not significantly different from the recovery RMSSD in the early and late samples, p = .860 and p = .814, respectively. In both the early and late samples, the effect of solution type and the interaction effect between solution type and interval type were not significant, with p = .322 and p = .509 for the early samples, and, p = .839 and p = .861 for the late samples, respectively.

3.3.1.2 Exploratory analysis 2: CRA word puzzle difficulty. Another analysis that might reveal differences between insight and non-insight is considering CRA word puzzle difficulty in relation to vmHRV. We observed that solution time and accuracy were closely linked, so that, on average, faster-solved puzzles were also more often solved correctly and vice versa. To confirm this observation, we first built an LMM, including solution time as the continuous outcome variable and solution accuracy as the binary predictor. As expected, we observed that correctly solved CRA word puzzles were solved faster (M = 7.73s) than incorrectly solved ones (M = 13.13s), $t(3098) = -19.08$, $p < .001$, Cohen's d = 0.91 (95% CI [0.82, 1.01]). This implies that solution time and accuracy jointly may inform us about the difficulty of an individual CRA word puzzle. CRA word puzzles that are, on average, solved faster and more often correctly can be considered the easier problems and vice versa for difficult ones. We applied two steps to define CRA word puzzle difficulty associated with insight and non-insight at the participant level. First, we calculated the Inverse Efficiency Score ($\text{IES} = \text{mean(solution time)}/\text{proportion correct}$; Vandierendonck, 2017) for each CRA word puzzle based on the participant's performance. The IES combines solution time and accuracy, where higher IES values are assumed to be related to a higher CRA word puzzle difficulty. Second, we calculated each participant's average IES values for their insightfully and non-insightfully solved CRA word puzzles, resulting in one average IES value for insight and one for non-insight per participant. Based on these values, we constructed a linear model with Box-Cox transformed baseline RMSSD as the outcome variable and solution type, IES values (standardized by rescaling them to z-scores), and their interaction term as predictors.

The results of the linear model showed a significant interaction effect between solution type and the scaled average IESs, $t(131) = 2.60$, $p = .010$ (see Table 5 in Appendix G for the model). For insight, a positive association between (Box-Cox transformed) RMSSD and the scaled average IES was observed, while a negative association between RMSSD and IES was observed for non-insight. This striking finding illustrates that participants with lower baseline RMSSD reported more often to have solved an easy problem with insight and a difficult problem with non-insight, whereas for participants with higher baseline RMSSD, the reverse was found. This result is illustrated in Figure 6. The main effects of solution type and scaled average IES were not significant, $p = .364$ and $p = .460$.

Figure 6

RMSSD Predicted by Solution Type and the Inverse Efficiency Score



Note. X-axis, scaled average inverse efficiency score (higher values = more difficult CRA word puzzles on average); Y-axis, RMSSD, root mean square of successive differences between normal inter-beat intervals; RMSSD was back-transformed from a Box-Cox transformation with $\lambda = 0.3$; red and green line represents the association between RMSSD and scaled average inverse efficiency score for non-insight and insight respectively; red and green shade, represents the 95% confidence intervals for the linear regression of non-insight and insight, respectively.

3.3.2 Assessing the Neurovisceral integration model

In the main analysis regarding trait vmHRV, we observed a negative association between solution accuracy and vmHRV and between the number of correctly solved word puzzles and vmHRV. This finding is inconsistent with the assumptions of the neurovisceral integration model, which would predict a positive association. **Exploratory analysis 2**, described above, also allowed us to further clarify this inconsistent observation by examining how RMSSD related to CRA word puzzle difficulty (i.e., the main effect of IES). In line with the neurovisceral integration model, we expected RMSSD to be positively associated with CRA word puzzle difficulty. However, the main effect of scaled average IES was not significant, $p = .460$ (see also exploratory analysis 2).

4. DISCUSSION

With the current study, we aimed to clarify how insight and non-insight are (in)differentially related to WMC from a trait and resource-dependent perspective. To that end, we asked participants to solve CRA word puzzles and measured vmHRV, an index of prefrontal cortex resources, at three different interval types: resting-state baseline, solution search, and recovery.

Our results showed that vmHRV measured during resting-state, indexing inter-individual differences in prefrontal functionality, was negatively associated with the problem-solving performance of both solution types (*trait*). Moreover, vmHRV reactivity during the solution search and its rebound during a post-task recovery period illustrated a similar pattern for both solution types (*resource-dependent*). Namely, vmHRV increased relative to its resting-state value during the solution search for both solution types. During the post-task recovery, vmHRV remained comparable to its value during the solution search. Psychophysiologically, we observed no differences between insight and non-insight problem solving, for the trait nor the state analyses. However, differences between insight and non-insight did emerge in exploratory analyses when splitting solution times in an early and late sample and when taking CRA word puzzle difficulty into account. Furthermore, behaviourally and metacognitively, our findings dissociated between the solution types, demonstrating that insight solutions were found faster, were more often correct, received higher solution confidence, and were solved more frequently than non-insightful ones.

4.1 Trait vagally mediated HRV

Opposing the direction of the hypothesized association between non-insight and vmHRV, we found that higher vmHRV, indexing inter-individual prefrontal functionality, was (marginally) associated with a decreased likelihood of solving CRA word puzzles correctly and a decrease in the frequency of the number of correctly solved CRA word puzzles. Importantly, this result was also found for insight, illustrating that both solution types relate to this intrinsic prefrontal functionality similarly. Splitting solution time in an early and late sample further elucidated this effect (exploratory analysis 1). Namely, this association between vmHRV and solution accuracy only reached significance in the late sample (i.e., observations > 7.006s). However, in the early sample (i.e., observations < 7.006s), the solution accuracies for insight (M = 96%) and non-insight (M = 89%) were perhaps close to ceiling level, leaving little room for variation in function of the participants' resting-state vmHRV, precluding the observation of a significant effect. This finding does show that the association is mainly driven by CRA word puzzles that required a longer solution time. Because there is, to our knowledge, no previous insight study that approached vmHRV from a trait perspective, it is difficult to make direct comparisons. Nonetheless, our findings align with some inter-individual studies on the association between WMC and insight and non-insight problem solving, showing that WMC is similarly involved for both solution types (Chein & Weisberg, 2014; Chuderski & Jastrzebski, 2018). On the other hand, our results are not in line with other studies, observing a differential involvement of WMC for both solution types (e.g., DeCaro et al., 2016; Gilhooly & Fioratou, 2009).

Thus, our results showed that higher inter-individual WMC/prefrontal functionality is associated with poorer CRA word puzzle problem solving performance. Some have argued that vmHRV primarily indexes inhibitory control instead of the full scope of executive functions (i.e., the ability to suppress unwanted thoughts and responses in service of the task goals; Kimhy et al., 2013; Munakata et al., 2011; Ottaviani et al., 2018). It might be that solving CRA word puzzles with insight or non-insight is hampered instead of aided by too much inhibitory control. Namely, with each step in the solution search, participants have to explore various solution candidates or alternative approaches to deal with the CRA word puzzle, even less obvious ones. To achieve this, the participant needs a loose associative mind, open to all types of information even though they may seem irrelevant at first glance, or to attend to some faint valuable information located at the border of consciousness. This type of thinking is also called divergent thinking (Zhang et al., 2020) and has been associated with CRA problem solving in general (Cancer et al., 2022; Wu & Chen, 2021) and CRA word puzzle solving with insight specifically (Jung-Beeman et al., 2004; Kounios & Jung-Beeman, 2015). If inhibitory control is too strong, filtering out seemingly irrelevant/less obvious associations or blocking information at the border of consciousness, this might hamper instead of aid CRA word puzzle solving. Indeed, previous studies have shown that divergent thinking benefits from less inhibitory control (e.g., Carson et al., 2003; Zabelina et al., 2016). Namely, less inhibitory control is presumed to facilitate the surfacing of more remotely associated task information (Abraham & Windmann, 2008; Cosgrave et al., 2018; but also see Nusbaum & Silvia, 2011). As such, those participants with a lower resting-state HRV, indexing lower prefrontal functionality (i.e., lower inhibitory control), might be more receptive to a broad range of ideas to proceed in the non-insight solution search or to access unusual approaches that can trigger insight. Although the conception of vmHRV indexing inhibitory control seems fruitful, we cannot confirm this interpretation solely based on our data as we did not include a behavioural measure of inhibitory control. Therefore, it would be worthwhile to assess inhibitory control alongside vmHRV in the context of problem solving to assess whether they are indeed interrelated, as suggested above. In any case, we speculate that our results indicate that the CRA word puzzle test might rely largely on divergent thinking, irrespective of how it is solved. Nevertheless, how this divergent thinking plays its role in insight and non-insight CRA word puzzle solving could still be different. It could be that vmHRV is too crude of a measure to pick up this dissociation in divergent thinking between the solution types. Therefore, it might be an exciting avenue of further inquiry to assess the nature of divergent thinking in both solution types and whether or not this is different by using a more fine-grained measure such as an electroencephalogram (EEG; e.g., Jia & Zeng, 2021).

Interestingly, when splitting solution time in an early and late sample (i.e., exploratory analysis 1), we observed for the early sample only, that the difference in confidence reported for insight and non-insight solutions became more pronounced with increasing resting-state vmHRV. One approach to clarify this, is by considering solution confidence as an index of metacognitive awareness — i.e., the propensity to be aware of one's ongoing thinking, which is closely associated with cognitive control, prefrontal functionality, and vmHRV (Allan et al., 2017; Ask et al., 2023, Fleur et al., 2021; Harrison & Vallin, 2018). Research has already shown that individuals with high vmHRV displayed better metacognitive awareness than their low vmHRV peers (e.g., Meessen et al., 2018; Ask et al., 2023). Therefore, it might be that higher resting-state vmHRV was associated with increased metacognitive awareness of the ongoing solution search, resulting in a differential judgment of insight and non-insight solutions in terms of confidence in the solution. Indeed, we observed that participants lower in resting-state vmHRV, relative to their higher vmHRV peers, judged both solution types more equally in terms of confidence, even though the insight and non-insight solutions were dissociable with regards to accuracy (i.e., 96% versus 89%). This shows that resting-state vmHRV was positively associated with the participants' metacognitive accuracy. However, this was only observed in the early sample (see Figure 6). The absence of the interaction effect in the late sample may result from a more pronounced difference in solution accuracy between insight and non-insight (i.e., 86% versus 57%) for these later solutions. This large solution accuracy difference was likely easier to track metacognitively, regardless of inter-individual differences in metacognitive awareness. One approach to further elucidate this interpretation is implementing a metacognitive awareness scale alongside measuring vmHRV and CRA word puzzles (see Harrison & Vallin, 2018) to assess the relationship more directly between confidence, metacognitive awareness, and vmHRV. Crucially, this result illustrates that not accounting for the variety in solution-time length in the original analysis obscured this differential relation between vmHRV and insight versus non-insight.

Furthermore, when taking into account CRA word puzzle difficulty (i.e., exploratory analysis 2), we observed a differential association between vmHRV and puzzle difficulty for insight and non-insight. Namely, individuals with lower resting-state vmHRV tended to judge easy CRA puzzles as solved with insight and difficult ones with non-insight, whereas the reverse was observed for higher resting-state vmHRV individuals. This observation, again, might be linked to divergent thinking. Divergent thinking involves two associative memory-search processes: spontaneous and goal-directed processes (Beaty & Kennet, 2023). The first entails accessing information spontaneously, like in a free association task (Merseal et al., 2023), and is linked to unconscious processing and neuronal activity in the default mode network (Marron et al., 2018). The second is a conscious process where individuals use executive control (e.g., inhibition and switching) to navigate memory, and is linked to neuronal activity in the executive control network (Beaty et al., 2014; Zhang et al., 2020). Certain studies, although scarce, indirectly imply that individuals with greater cognitive resources are biased towards goal-directed processing, while those with fewer resources are biased towards more spontaneous processing (e.g., Robison et al., 2020; Liu et al., 2023). Therefore, we tentatively argue that the preference for one over the other memory-search process might depend on an individual’s prefrontal functionality (i.e., as indexed by vmHRV). Specifically, individuals with lower resting-state vmHRV, often associated with reduced cognitive resources, might rely more on a spontaneous search because cognitive control engagement would place additional strain on their already constrained cognitive resources. Conversely, those with higher resting-state vmHRV, often associated with enhanced cognitive resources, might favour a goal-directed search, exploiting the benefit of their greater cognitive resources. This propensity, in turn, could impact how easy and difficult CRA puzzles are experienced in terms of insight and non-insight.

Specifically, for **easy CRA puzzles**, lower vmHRV individuals relying more on spontaneous search, likely access the solution spontaneously after only one or two attempts, resulting in a sudden insight experience (see also Becker et al., 2021). Contrarily, higher vmHRV individuals relying more on goal-directed search, might perceive easy CRA puzzles as non-insightful because they can navigate the semantic search space in a goal-directed way, facilitated by the limited scope of the solution-space for these easy puzzles. Our results with regards to the early sample's solution confidence support this, highlighting an increased awareness of the ongoing solution search in higher vmHRV individuals, aligning with goal-directed search processes. On the other hand, for **difficult CRA puzzles**, lower vmHRV individuals may need multiple cycles of spontaneous search or resort to goal-directed search after spontaneous attempts fail, leading to the experience of non-insight. Contrarily, higher vmHRV individuals might encounter dead ends in the semantic search space, necessitating the inhibition of erroneous assumptions (cf. restructuring) or the reliance on a spontaneous search to find the solution. Such a sudden shift in the protracted solution search might be experienced as an insight. In any case, an individual's tendency to rely more on spontaneous or goal-directed associative searches of semantic memory might influence how they experience and interpret the process of solving CRA word puzzles, particularly in terms of insightfulness. However, as we only indirectly reached these conclusions after an exploratory analysis, we want to stress that caution is needed here and future research is required to test these hypotheses more directly. One possible way to assess this more directly is by examining the relative involvement of the default mode and executive control network during CRA solving of individuals high and low in cognitive resources (e.g., based on cognitive control test battery), while also tracking vmHRV.

Finally, the results of the main analyses were difficult to reconcile with the neurovisceral integration model's assumption that individuals with higher vmHRV tend to have better prefrontal resources for self-regulatory purposes than their lower vmHRV peers (Thayer et al., 2009). Namely, instead of observing a positive association between vmHRV and behavioural CRA performance, a negative one was found. To elucidate this unexpected association, we additionally examined CRA word puzzle difficulty in association with vmHRV, irrespective of solution type, using exploratory analysis 2. CRA word puzzle difficulty appeared unrelated to vmHRV. Although these findings are not in line with the neurovisceral integration model, the findings of exploratory analyses 1 (early and late solution retrieval) and 2 (CRA word puzzle difficulty) tend to align with this model. Namely, these analyses showed that higher vmHRV individuals are more aware of their ongoing solution search and are arguably more prone to engage in a goal-directed search, characteristics which the model would associate with higher vmHRV.

In conclusion, although the main analyses showed a similar association between vmHRV and behavioural CRA performance for both solution types, the more fine-grained exploratory analyses revealed that insight and non-insight were dissociable when considering the length of the solution search and CRA word puzzle difficulty. While an enhanced inhibitory control of higher vmHRV individuals may hamper CRA problem solving, they might, at the same time, also be more aware of the ongoing solution search and adopt a more goal-directed approach to solve CRA puzzles.

4.2 Resource-dependent HRV

We observed that vmHRV increased from resting-state baseline to solution search for both solution types. Moreover, this vmHRV remained at a comparable level during the post-task recovery period. The exploratory analysis revealed a similar result, where solution time was split into an early and late sample. Taken that the solution search interval was not divided into its constituent solution search phases (i.e., problem representation, search, and solution retrieval), these findings show that it is valuable to consider the full solution search to assess the involvement of prefrontal resources. However, it might have been that the different solution search phases displayed a differential prefrontal involvement for insight and non-insight. This would require dividing each CRA word puzzle trial into its solution search phases. The question then arises of how to determine the time interval of each solution search phase. This would not be trivial. First, inter-individual differences might occur in the time spent in each solution search phase, with some individuals devoting more time to certain phases than others. Second, it is likely (given the differences in solution time between the trials, i.e., range 1.26s – 29.68s) that the intervals for each phase in the solution search vary from trial to trial. Furthermore, if these phases are very short, this would hamper the assessment of vmHRV. For example, regarding solution retrieval, previous psychophysiological studies (e.g., Jung-Beeman et al., 2004; Salvi et al., 2020) have shown differential patterns in EEG and pupil dilation data for insight and non-insight in the last 500ms before solution retrieval. If this 500ms interval represents the solution retrieval phase in the CRA-task context, using vmHRV to index prefrontal resources for insight and non-insight would be impossible. Therefore, having an accurate demarcation of each solution search phase for each trial would be interesting. Perhaps this can be based on participants' online subjective experiences during the solution search (e.g., Fedor et al., 2015) and/or by tracking changes in eye behaviour (e.g., Salvi et al., 2020). In any case, our results show that when we take into account the full solution search, there is something markedly similar between the two solution types.

Namely, our findings highlighted that both solution types relate to the use of prefrontal resources in a similar vein. This observation is not in line with Jausovec and Bakracevic's (1995) study, which found a differential pattern for both solution types. Studying heart rate, they found an incremental increase, relative to baseline, for non-insight (measured via verbal math problems) and a sudden increase at the moment of solution retrieval for insight (measured via visuospatial riddles). For non-insight, one would expect this increase in heart rate to correspond to a vmHRV decrease, similar to what is seen in studies assessing vmHRV reactivity while solving math-like problems (Singh et al., 2019; Sloan et al., 1991). For insight, the expected pattern of the vmHRV reactivity (i.e., increased or unaffected relative to its resting-state baseline) is less clear based on their heart rate study. In any case, it is surprising that the non-insightful solution search in the present study, which we considered to depend on the exertion of mental effort, was not associated with the depletion of prefrontal resources which would have been indexed by a decrease in vmHRV.

vmHRV decreases are most often found during tasks for which demands are high, but it is clear how they should be procedurally executed (e.g., WM task, math problems; Hansen et al., 2003; Overbeek et al., 2014; Singh et al., 2019). However, the nature of such tasks might differ greatly from the CRA word puzzle test used in the current study. Namely, figuring out the solution to the CRA word puzzles is not an everyday task for which procedural knowledge is present. So solving the CRA word puzzles might require a different approach that taxes prefrontal resources differently than, for instance, solving a math problem. As the CRA test is based on the remote associates test of Mednick (1962), a well-known test to assess creativity, it is not inconceivable that reaching a solution (non)insightfully is, in a larger part, a creative challenge (Wu & Chen, 2021) rather than the application of procedural knowledge. Indeed, not all HRV research has illustrated a task-dependent decrease in vmHRV (e.g., Silvia et al., 2014). For example, it has been found that task contexts relating to impulse control (e.g., resisting the urge to drink alcohol) and creativity are found to be associated with vmHRV *increases* (Denson et al., 2011; Ingjaldsson et al., 2003; Segerstrom & Nes, 2007). One of the arguments is that increasing prefrontal resource, as indexed by elevated vmHRV, during such tasks is adaptive here as it promotes calm reflection to approach a situation from a neutral and open stance (Denson et al., 2011; Laborde et al., 2018; Rominger et al., 2019; Segerstrom & Nes, 2007). Such an open-minded, calm reflection might support divergent thinking, which is often associated with creativity (see Cancer et al., 2022; Zhang et al., 2020). This observation neatly aligns with our results concerning the trait vmHRV described in the previous section. Namely, solving CRA word puzzles might be mainly a creative challenge depending on divergent thinking, regardless of whether the puzzle is solved with insight or non-insight.

It has recently been argued that creative problem-solving and insight problem-solving specifically rely on divergent/flexible thinking and convergent/persistent thinking (Zhang et al., 2020). The latter is used to narrow down and evaluate the options to continue the solution search or attain the solution (Cancer et al., 2022; Zhang et al., 2020). One or multiple cycles of divergent and convergent thinking might be needed to find the solution (Hélie & Sun, 2010; Kajic et al., 2017; Rominger et al., 2019). Insight might be preceded by only one or two such cycles to restructure the erroneous solution space and attain the solution, perhaps still relying on implicit processes to a large extent. In contrast, non-insight might depend on multiple cycles building on previous cycles through the solution search. Becker et al. (2021) showed that more solution attempts precede non-insight than insight, tentatively corroborating this assumption. Thus, future research pinpointing whether these two modes of thinking are present in insight and non-insight problem solving and whether these modes differ between insight and non-insight in how they fluctuate temporally would be worthwhile.

The finding that vmHRV did not progress back to its resting-state value during the recovery period might not be that surprising. The body's preferred state is one that enhances resource restoration and rest, marked by parasympathetic dominance and, thus, increased vmHRV (Laborde et al., 2017; Shaffer & Ginsberg, 2017). Therefore, maintaining a higher vmHRV post-task can be seen as adaptive, reflecting that prefrontal resources remain more readily accessible to address the next CRA word puzzle trial (see also Laborde et al., 2017). It is also not inconceivable that the experience of insight, compared to non-insight, induces vmHRV changes due to insight's surprising and affective nature. Although this would be an exciting avenue to further establish the psychophysiological difference between insight and non-insight, the current experimental design does not allow us to test this hypothesis. Namely, participants' typing behaviour immediately following solution retrieval confounds any immediate post-solution vmHRV changes related to solution type.

Finally, it is noteworthy that we merged all correctly solved CRA word puzzle trials with insight and non-insight to obtain valid measures of vmHRV. This merging into large solution search and recovery intervals might not have done justice to the diversity of solution search strategies present within each CRA word puzzle trial. Perhaps, as a next step, it would be interesting to use more fine-grained measures, such as EEG and eye-tracking, to tease apart the insightful versus non-insightful solution search on a more trial-by-trial basis (see, for example, Becker et al., 2021).

4.3 Behavioural and metacognitive differences

Our behavioural and metacognitive results for each solution type correspond to the findings of numerous other studies (e.g., Danek & Wiley, 2017; Hedne et al., 2016; Laukkonen et al., 2021; Salvi et al., 2016; Stuyck et al., 2022; Webb et al., 2019). That is, insight solutions are typically solved faster, are more often correct, receive more confidence, and are solved more frequently than non-insightful ones. Only solution time has been considered a less stable characteristic separating insight from non-insight, with some studies finding faster solution times for insight (e.g., Cranford & Moss, 2012), no difference (e.g., Hedne et al., 2016), or slower solution times (e.g., Stuyck et al., 2021). Nonetheless, these observations ensure that the current study's self-reported insight and non-insight solution strategies closely mimic what has been previously observed in insight studies using the same paradigm. It is worth mentioning that these self-reports have not been without critique. For instance, it has been claimed that insight classification is driven by finding a solution quickly, therefore bypassing an actual solution search and artificially increasing the accuracy effect (i.e., fast = easy = more accurate; Cranford & Moss; Stuyck et al., 2022). However, several studies have already shown that accounting for these fast insight solutions does not change the observed patterns (e.g., Salvi et al., 2016; Stuyck et al., 2021; Stuyck et al., 2022). Moreover, it has been argued that the definition of insight provided to participants at the onset of the experiment and/or participants' intrinsic conception of insight biases participants to rely on certain phenomenological (e.g., solution confidence) and behavioural (e.g., solution accuracy) cues to make the insight/non-insight classification (Laukkonen et al., 2021; Laukkonen & Tangen, 2018). However, research has shown that the insight/non-insight classification corresponds to a distinct psychophysiological (e.g., squeeze strength, pupil dilation; Laukkonen et al., 2021; Salvi et al., 2020), neurological (e.g., EEG signature; see Kounios & Jung-Beeman, 2014 for a review), and behavioural (e.g., differentially affected by cognitive load; Stuyck et al., 2022) signature, thereby illustrating that the self-reports implicate differing constructs. Notably, we observed that some participants proportionally solved more word puzzles with insight than non-insight ($N = 17$) or vice versa ($N = 14$). This might reflect inter-individual differences in the propensity to solve problems with insight or non-insight. Therefore, it might be an interesting avenue to explore this insight/non-insight propensity in future research, ideally using a longitudinal design to examine whether the insight/non-insight propensity is a robust phenomenon over time.

Furthermore, we must note that the current sample consisted almost exclusively of female undergraduates. Previous research has not found an effect of biological sex on the tendency to solve problems with insight or non-insight (e.g., Stuyck et al., 2022; Wieth & Burns, 2006). On the other hand, the effect of biological sex on HRV has been somewhat inconsistent with some studies (e.g., Alyahya et al., 2021) showing no difference in young adults and others reporting otherwise (e.g., Estévez-Baés et al., 2018). In any case, our results should be interpreted with the composition characteristics of the sample in mind. Lastly, due to COVID-19, we had to adhere to strict ethical guidelines. To avoid any undue influence on HRV measurements, participants did not wear face masks while being tested. However, participants were requested to attach the electrodes themselves. Although we provided verbal and pictorial guidance and visually double-checked the electrode attachment, we cannot exclude the possibility that, on some occasions, the electrode placement was suboptimal, leading to poor ECG data. This could explain the higher proportion of excluded participants than usual. When no health-related restrictions or other ethical issues apply, we recommend that an experimenter attaches the electrodes to ensure high-quality data.

4.4 Conclusion

Trait vmHRV was negatively associated with CRA problem solving performance. We argued that this might reflect inter-individual differences in inhibitory control. As the solution search requires one to think of remote associations, inhibitory control might hamper rather than aid this process. This was further substantiated by observing a vmHRV increase from resting-state baseline to solution search, which lingered on in the recovery period. This vmHRV increase could mark the increase of prefrontal resources to promote an open-minded stance, important for thinking in a divergent manner, which arguably is crucial for the CRA test used in this study. Although this psychophysiological pattern of results was observed for both solution types, we also found that metacognitively differentiating insight from non-insight based on confidence in the early sample was positively associated with trait vmHRV (exploratory analysis 1). We suggested that higher trait vmHRV individuals have more fine-grained metacognitive information about the ongoing solution search, enabling them to guide their solution confidence effectively. Furthermore, trait vmHRV was positively associated with CRA puzzle difficulty for insight and negatively associated with CRA puzzle difficulty for non-insight (exploratory analysis 2). We proposed that this observation might stem from higher trait vmHRV individuals favouring goal-directed divergent thinking while lower vmHRV individuals might favour spontaneous divergent thinking, resulting in a differential judgment of whether easy and difficult CRA puzzles were solved using insight or non-insight. Although the main analyses indicated that insight and non-insight shared a similar association between vmHRV and behavioral performance, the more fine-grained exploratory analyses revealed that insight and non-insight were metacognitively judged differently in the early sample of CRA word puzzles and were experienced differently depending on the CRA word puzzle difficulty.

ACKNOWLEDGMENT

We thank Yujing Liang and Amar Music for their assistance with the data collection and their critical thought and the Fonds de la Recherche Scientifique and the Research Foundation Flanders (FWO) 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] and the Research Foundation Flanders [grant number G096919N]. Christo Bratanov received an international mobility research grant from the Institut Servier.

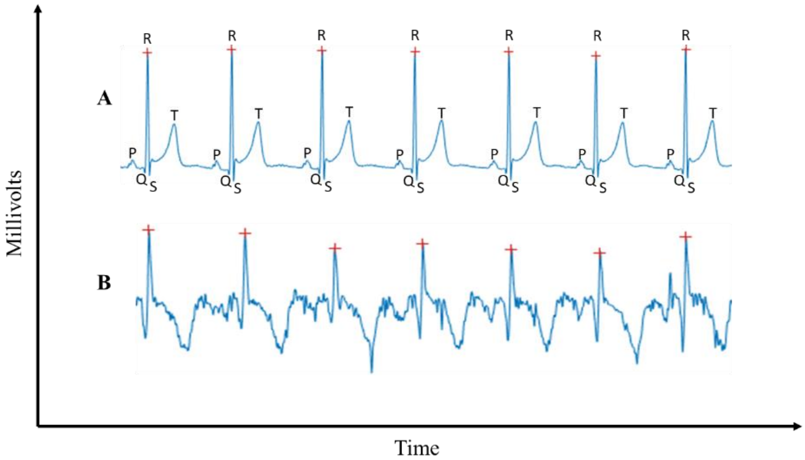
DECLARATION OF CONFLICTING INTERESTS

The Authors declare that there is no conflict of interest

Appendix A. Example of abnormal ECG signal

Figure 1

Examples of a Normal and Abnormal ECG signal



Note. A, normal ECG signal; B, abnormal ECG signal; P, P-wave (depolarization of the atria); Q, Q-wave (depolarization of the interventricular septum); R, R-wave (depolarization of the main body of the ventricles); S, S-wave (depolarization of the ventricles at the base of the heart); T, T-wave (repolarization of the ventricles; Silverthorn, 2004, p. 448); the time interval between two consecutive R-waves is used to estimate the IBIs.

Appendix B. Instructions

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 30 s to find a solution. Once you have found the solution, press the space bar and enter your answer.

In addition, you must indicate whether you have solved the word 'with an Aha!' or 'without an 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!".

Finally you should also 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".

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.

Appendix C. Data preprocessing

To synchronize the course of the experiment programmed in PsychoPy with the ECG signal, we used the Nexus trigger interface to send triggers from PsychoPy to the ECG recording software (i.e., BioTrace+), marking the intervals of interest (i.e., baseline, solution search, and recovery).

2.6.1 Kubios. To detect the r-peaks in the ECG signal, the Kubios HRV software uses a QRS-detection algorithm based on the Pan-Tompkins algorithm (see Tarvainen et al., 2020, for an in-depth explanation). After that, the IBIs were calculated by determining the time interval between two r-peaks for the ECG signal (see Fig. 2 for an example). To correct potential artifacts in the IBI time series, we used the automatic artifact correction algorithm of Kubios. This algorithm uses the distribution of the differences between consecutive IBIs to estimate a time-varying threshold (i.e., threshold changes depending on location in the IBI time series) to identify ectopic beats. To detect missed and extra beats, the time-varying threshold is determined based on the distribution of the differences between the IBIs and a local, median IBI (see Lipponen & Tarvainen, 2019, for the algorithm and decision rule). All detected artifacts are subsequently replaced with IBIs based on cubic spline interpolation. Lastly, Kubios deploys a detrending procedure to accommodate the non-stationarity of the IBI time series by defining an a priori smoothing parameter (cut-off frequency 0.035 Hz; see Tarvainen et al., 2002).

2.6.2 Visual inspection. Additionally, all ECG signals were visually inspected for abnormal ECG signals, unstable recording epochs, missed r-peaks, and missed artifacts by the algorithms that might influence the vmHRV data (e.g., supraventricular extrasystole; see Kumral et al., 2019 for a similar procedure). In case of abnormalities, we applied a manual correction to the ECG signal (e.g., marking noisy ECG epochs as to-be-excluded noise epochs and/or adding missing r-peaks). For all ECG recording intervals (i.e., baseline, solution search, and recovery), we only accepted ECG recordings consisting of at least 95% noise-free data (i.e., clear and distinct ECG wave morphology) and no more than 5% corrected IBIs in that noise-free data. For example, a baseline ECG recording of 300s could not include more than 15s of noise epochs (i.e., an epoch with an uninterpretable ECG wave morphology).

2.6.3 Baseline ECG recording. As the 5min baseline ECG recording is vital for the statistical analysis of the trait vmHRV and the resource-dependent vmHRV, baseline ECG recordings that violated these above percentages (i.e., min. 95% noise-free data and max. 5% corrected IBIs) led to the exclusion of one participant (see participant section). Table 1 depicts the percentage of corrected IBIs and noise-free data.

2.6.4 Solution search and recovery ECG recording. To create the *solution search* and *recovery* intervals over which RMSSD was calculated, we merged the time intervals of the individually correctly solved CRA word puzzle trials for each solution type. This led to four RMSSD observations per participant (i.e., solution search insight, solution search non-insight, recovery insight, and recovery non-insight). The merged *solution-search* interval consisted of a series of time intervals with differing lengths. This is because CRA word puzzles took between 1s and 30s to be solved. The merged *recovery* interval always consisted of a series of time intervals of 10s, as the recovery time interval after each trial had a fixed length of 10s. The minimum *solution search* and *recovery* interval length deemed acceptable for assessing RMSSD was 10s. Previous research (e.g., Munoz et al., 2015) has shown that RMSSD calculated with a 10s ECG recording gives a reliable approximation of the RMSSD obtained with a 5min ECG recording. Based on this minimum required interval length, we refrained from calculating RMSSD for the insightful *solution search* of one participant. Next, similar to the baseline ECG recording, we determined the percentage of noise-free data and the percentage of corrected IBIs in the noise-free data. If these percentages were below 95% and above 5%, respectively, we refrained from calculating the RMSSD value but retained the participant. This led to an additional omission of five RMSSD values of five different participants. We note that for the *solution search* and *recovery* intervals below 30s all ECG data were 100% noise-free without corrected IBIs. This is important, as for such short intervals with a low number of IBIs which can be used to calculate RMSSD, additional missing/erroneous ECG data would be detrimental to the valid estimation of RMSSD. Finally, we used Tukey's (1977) method (see participant section for explanation) to identify severely outlying RMSSD observations in the four different interval types (i.e., solution search insight, solution search non-insight, recovery insight, and recovery non-insight; see Kumral et al., 2019 for a similar procedure). No RMSSD observations were considered as outlying based on this method. Table 1 depicts the percentage of noise-free data, the percentage of corrected IBIs, and the number of excluded RMSSD observations for each interval type.

Table 1

The Percentage of IBIs corrected and the Percentage of Noise-Free Data

		%corrected IBI		%noise-free ECG data		#RMSSD excluded
		M(*SD*)	range	*M*(*SD*)	range	
	baseline	0.26(0.49)	0-1.82	99.88(0.63)	95.32-100	/
Insight	solution search	0.10(0.35)	0-2.34	99.75(2.06)	83.00-100	2
	recovery	0.36(0.74)	0-4.15	99.88(0.66)	95.00-100	/
Non-Insight	solution search	0.38(0.99)	0-5.26	99.96(0.37)	97.00-100	1
	recovery	0.63(1.40)	0-6.06	99.97(0.24)	98.00-100	3

Note. %corrected IBI, percentage of IBIs that were corrected in the noise-free data; %noise-free ECG data, percentage of data with a clear and distinct ECG signal; RMSSD, root mean square of successive differences between normal IBIs; #RMSSD excluded, number of RMSSD observations within a specific interval type that were excluded because they either had too much IBIs corrected or had insufficient noise-free ECG data; /, no excluded RMSSD observations.

Appendix D. R package used for the statistical analysis

The (G)LMMs of trait vmHRV of solution time and solution accuracy, and the LMM of resource-dependent vmHRV of RMSSD were built with the lme4 package (Bates et al., 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). Box-Cox transformations were applied with the MASS package (Venable & Ripley, 2002). Post-hoc tests were performed with the emmeans package (Lenth, 2020).

Appendix E. Trait vagally mediated HRV's estimated models

Table 1.

Linear Mixed Model on Solution Time of the CRA

	β(*SE*)	CI 95%	*t*-value	*p*
Intercept (grand mean)	1.69(0.03)	1.64, 1.74	67.66	**<.001**
solution type	-0.16(0.02)	-0.19, -0.13	-9.31	**<.001**
RMSSD	-0.00001(0.02)	-0.03, 0.04	-0.001	.999
solution type*RMSSD	-0.019(0.02)	-0.06, 0.02	-1.14	.256

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *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 CRA word puzzles.

Table 2.

Generalized Linear Mixed Model on Solution Accuracy of the CRA

	β(*SE*)	*OR*	CI 95%	*Z*-value	*p*
Intercept (grand mean)	1.62(0.16)	5.07	3.75, 6.87	10.51	**<.001**
solution type	1.74(0.12)	5.67	4.53, 7.11	15.07	**<.001**
RMSSD	-0.22(0.12)	0.81	0.64, 1.01	-1.85	**.065**
solution type*RMSSD	-0.09(0.11)	0.92	0.74, 1.14	-0.79	.436

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *p*-values were based on the Wald test; Betas are on the logit scale; *OR,* odds ratio; an *OR* of one represents the at chance-level classification of correct and incorrectly solved CRA word puzzles; An *OR* above/below one represents the magnitude of increase/decrease in the probability of solving a CRA word puzzle correctly; Boldface, significant and borderline significant results; CI, confidence interval. This generalized linear mixed model is based on the correct and incorrect solved CRA word puzzles.

Table 3.

Generalized Linear Mixed Model on Solution Confidence of the CRA

	β(*SE*)	CI 95%	*Z*-value	*p*
Intercept (grand mean)	1.11(0.10)	0.93, 1.30	11.75	**<.001**
solution type	0.62(0.04)	0.54, 0.70	14.42	**<.001**
RMSSD	0.06(0.08)	-0.10, 0.21	0.72	.474
solution type*RMSSD	0.03(0.04)	-0.06, 0.11	0.68	.496

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *p*-values were based on the Wald test; Betas are on the logit scale; Boldface, significant results; CI, confidence interval. This generalized linear mixed model is based only on the correct solved CRA word puzzles.

Table 4.

Generalized Linear Mixed Model on the Number of Correctly Solved CRA Word Puzzles

	β(*SE*)	*IRR*	CI 95%	*Z*-value	*p*
Intercept (grand mean)	2.85(0.03)	17.23	15.90, 18.68	90.55	**<.001**
solution type	0.55(0.09)	1.73	1.47, 2.03	5.97	**<.001**
RMSSD	-0.06(0.03)	0.94	0.87, 1.02	-2.26	**.024**
solution type*RMSSD	0.13(0.10)	1.14	0.97, 1.33	1.34	.181

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *p*-values were based on the Wald test; Betas are on the log scale; *IRR,* incidence rate ratio; an *IRR* of one represents no change in the rate of the number of correctly solved CRA word puzzles; An *IRR* above/below one represents the multiplicative factor of increase/decrease in the rate of the number of correctly solved CRA word puzzles; Robust standard errors were computed to take the non-independence of the data into account; Boldface, significant results; CI, confidence interval. This generalized linear model is based on only the correctly solved CRA word puzzles.

Appendix F. Resource-dependent vagally mediated HRV

Table 1.
Linear Mixed Model on (Box-Cox transformed) RMSSD

	β(*SE*)	CI 95%	*t*-value	*p*
Intercept (grand mean)	6.63(0.13)	6.38, 6.91	51.85	**<.001**
solution type	-0.01(0.05)	-0.11, 0.10	-0.23	.822
search interval1	0.32(0.08)	0.18, 0.48	4.13	**<.001**
search interval2	-0.43(0.08)	-0.58, -0.29	-5.66	**<.001**
solution type*search interval1	0.13(0.15)	-0.17, 0.41	0.84	.402
solution type*search interval2	0.15(0.15)	-0.16, 0.49	1.00	.319

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals; RMSSD is Box-Cox transformed with λ = 0.3 to accommodate non-normality and heteroscedasticity; search interval1, baseline vs. solution search; search interval2, baseline vs. recovery; *p*-values were obtained using the Satterthwaite approximation method; Boldface, significant results; CI, confidence interval; this linear mixed model is based on the RMSSD calculated on the merged intervals of the solution search and recovery intervals for the correctly solved CRA word puzzles.

Appendix G. Exploratory analysis

Table 1.
Linear Mixed Model on Solution Time of the CRA for the Early and Late Samples

	Early			
	β(*SE*)	CI 95%	*t*-value	*p*
Intercept (grand mean)	2.96(0.05)	2.86, 3.08	99.22	**<.001**
solution type	-0.21(0.06)	-0.32, -0.12	-3.65	**<.001**
RMSSD	0.03(0.04)	-0.07, 0.10	0.61	.544
solution type*RMSSD	-0.09(0.06)	-0.19, 0.03	-1.49	.136
	Late			
	β(*SE*)	CI 95%	*t*-value	*p*
Intercept (grand mean)	1.48(0.005)	1.47, 1.49	322.31	**<.001**
solution type	-0.03(0.007)	-0.04, -0.009	-3.54	**<.001**
RMSSD	-0.006(0.004)	-0.01, 0.004	-1.12	.266
solution type*RMSSD	-0.0004(0.0004)	-0.02, 0.02	-0.06	.954

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *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 CRA word puzzles in the early (observations < 7s) and the late sample (observations > 7s).

Table 2.
Generalized Linear Mixed Model on Solution Accuracy of the CRA for the Early and Late Samples

	Early				
	β(*SE*)	*OR*	CI 95%	*t*-value	*p*
Intercept (grand mean)	2.70(0.23)	14.91	9.45, 23.53	11.61	**<.001**
solution type	1.17(0.22)	3.21	2.07, 4.96	5.26	**<.001**
RMSSD	-0.22(0.16)	0.81	0.59, 1.11	-1.31	.191
solution type*RMSSD	0.32(0.22)	1.37	0.89, 2.13	1.42	.156
	Late				
	β(*SE*)	*OR*	CI 95%	*t*-value	*p*
Intercept (grand mean)	1.07(0.14)	2.88	2.18, 3.80	7.46	**<.001**
solution type	1.58(0.15)	4.83	3.64, 6.42	10.88	**<.001**
RMSSD	-0.24(0.11)	0.79	0.63, 0.98	-2.10	**.036**
solution type*RMSSD	-0.20(0.14)	0.82	0.62, 1.08	-1.44	.149

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *p*-values were based on the Wald test; Betas are on the logit scale; *OR*, odds ratio; an *OR* of one represents the at chance-level classification of correct and incorrectly solved CRA word puzzles; An *OR* above/below one represents the magnitude of increase/decrease in the probability of solving a CRA word puzzle correctly; Boldface, significant and borderline significant results; CI, confidence interval. This generalized linear mixed model is based on the correct and incorrect solved CRA word puzzles in the early (observations < 7s) and the late sample (observations > 7s).

Table 3.

Generalized Linear Mixed Model on Solution Confidence of the CRA for the Early and Late Samples

	β(*SE*)	CI 95%	*t*-value	*p*
Early				
Intercept (grand mean)	1.57(0.10)	1.38, 1.76	16.07	**<.001**
solution type	0.40(0.07)	0.28, 0.53	6.20	**<.001**
RMSSD	0.04(0.09)	-0.13, 0.21	0.50	.619
solution type*RMSSD	0.14(0.06)	0.02, 0.27	2.19	**.028**
Late				
Intercept (grand mean)	0.91(0.10)	0.73, 1.10	9.62	**<.001**
solution type	0.72(0.06)	0.60, 0.84	11.80	**<.001**
RMSSD	0.06(0.08)	-0.10, 0.21	0.71	.475
solution type*RMSSD	-0.02(0.06)	-0.14, 0.10	-0.28	.783

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals standardized by rescaling it to z-scores; *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 CRA word puzzles in the early (observations < 7s) and the late sample (observations > 7s).

Table 4.

Linear Mixed Model on (Box-Cox transformed) RMSSD for the Early and Late Samples

	β(*SE*)	CI 95%	*t*-value	*p*
Early				
Intercept (grand mean)	5.33(0.09)	5.16, 5.50	60.04	**<.001**
solution type	0.05(0.05)	-0.04, 0.13	0.99	.322
search interval1	0.26(0.06)	0.15, 0.38	4.14	**<.001**
search interval2	-0.20(0.07)	-0.33, -0.06	-3.05	**.003**
solution type*search interval1	-0.06(0.13)	-0.34, 0.17	-0.50	.620
solution type*search interval2	0.15(0.13)	-0.47, 0.08	-1.15	.250
Late				
Intercept (grand mean)	5.33(0.09)	5.15, 5.50	58.99	**<.001**
solution type	0.009(0.04)	-0.07, 0.08	0.20	.839
search interval1	0.25(0.06)	0.15, 0.37	4.20	**<.001**
search interval2	-0.19(0.06)	-0.30, -0.08	-3.17	**.002**
solution type*search interval1	-0.05(0.12)	-0.31, 0.20	-0.38	.701
solution type*search interval2	-0.06(0.12)	-0.30, 0.19	-0.53	.600

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals; RMSSD is Box-Cox transformed with $\lambda = 0.3$ to accommodate non-normality and heteroscedasticity; search interval1, baseline vs. solution search; search interval2, baseline vs. recovery; *p*-values were obtained using the Satterthwaite approximation method; Boldface, significant results; CI, confidence interval; this linear mixed model is based on the RMSSD calculated on the merged intervals of the solution search and recovery intervals for the correctly solved CRA word puzzles in the early (observations < 7s) and the late sample (observations > 7s).

Table 5.

Linear Model on (Box-Cox transformed) RMSSD

	β(*SE*)	CI 95%	*t*-value	*p*
Intercept (grand mean)	6.39(0.13)	6.14, 6.64	50.19	**<.001**
solution type	-0.07(0.08)	-0.23, 0.09	-0.91	.364
scaled average IES	-0.07(0.09)	-0.25, 0.12	-0.74	.460
solution type*scaled average IES	0.59(0.23)	0.14, 1.03	2.60	**.010**

Note. RMSSD, root mean square of successive differences between normal inter-beat intervals; RMSSD is Box-Cox transformed with $\lambda = 0.3$ to accommodate non-normality and heteroscedasticity; scaled average IES, average inverse efficiency score for insight and non-insight per participant standardized by rescaling it to z-scores; Boldface, significant results; CI, confidence interval. This linear model is based on only the correctly solved CRA word puzzles.

REFERENCES

Abraham, A., & Windmann, S. (2008). Selective information processing advantages in creative cognition as a function of schizotypy. *Creativity Research Journal*, *20*(1), 1–6. doi: 10.1080/10400410701839819

Allan, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., Callaghan, M. F., & Rees, G. (2017). Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage*, *149*(April), 415–423. doi.org/10.1016/j.neuroimage.2017.02.008

Alyahya, A., Fuller, A., Okwose, N., Charman, S., Macgowan, G., & Jakovljevic, D. (2021).The effect of age and gender on heart rate variability in healthy individuals. *European Journal of Preventive Cardiology*, *28*(Supplement_1). doi.org/10.1093/eurjpc/zwab061.255

Ask, T. F., Knox, B. J., Lugo, R. G., Helgetun, I., & Sütterlin, S. (2023). Neurophysiological and emotional influences on team communication and metacognitive cyber situational awareness during a cyber engineering exercise. *Frontiers in Human Neuroscience*, *16*(January), 1–24. doi:10.3389/fnhum.2022.1092056

Atkinson, A. C., Riani, M., & Corbellini, A. (2021). The Box–Cox transformation: Review and extensions. *Statistical Science, 36*(2), 239–255. doi: 10.1214/20-STS778

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. doi: 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. doi: 10.21500/20112084.807

Balzarotti, S., Biassoni, F., Colombo, B., & Ciceri, M. R. (2017). Cardiac vagal control as a marker of emotion regulation in healthy adults: A review. *Biological Psychology, 130,* 54–66. doi: 10.1016/j.biopsycho.2017.10.008

Bates, Douglas, Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–29. doi: 10.18637/jss.v067.i01

Beaty, R. E., Benedek, M., Wilkins, R. W., Jauk, E., Fink, A., Silvia, P. J., Hodges, D. A., Koschutnig, K., & Neubauer, A. C. (2014). Creativity and the default mode network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia, 64*(September), 92–98. dx.doi.org/10.1016/j.neuropsychologia.2014.09.019

Beaty, R. E., & Kennet, N. K. (2023). Associative thinking at the core of creativity. *Trends in Cognitive Science*, *27*(7), 671–683. doi.org/10.1016/j.tics.2023.04.004

Becker, M., Kühn, S., & Sommer, T. (2021). Verbal insight revisited—Dissociable neurocognitive processes underlying solutions accompanied by an AHA! experience with and without prior restructuring. *Journal of Cognitive Psychology*, *33*, 659–684. doi: 10.1080/20445911.2020.1819297

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*, *27*, 116-122. doi: 10.1016/j.cobeha.2018.11.002

Bornstein, M. H., & Suess, P. E. (2000). Child and mother cardiac vagal tone: Continuity, stability, and concordance across the first 5 years. *Developmental Psychology*, *36*(1), 54–65. doi: 10.1037/0012-1649.36.1.54

Bowden, E. M., & Grunewald, K. (2018). Whose insight is it anyway?. *Insight: On the Orgins of New Ideas* (pp. 28–50). In F. Vallée-Tourangeau (Ed.).

Bowden, E. M., & Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods*, *42*(2007), 87–99. doi: 10.1016/j.ymeth.2006.11.007

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & 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. doi: 10.3929/ethz-b-000240890

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A Tutorial. *Journal of Cognition*, *1*(1), 1–20. doi: 10.5334/joc.10

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. doi: 10.3758/s13428-012-0186-0

Cancer, A., Iannello, P., Salvi, C., & Antonietti, A. (2022). Executive functioning and divergent thinking predict creative problem-solving in young adults and elderlies. *Psychological Research*. https://doi.org/10.1007/s00426-022-01678-8

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2003). Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals. *Journal of Personality and Social Psychology*, *85*(3), 499–506. doi: 10.1037/0022-3514.85.3.499

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. doi: 10.3758/s13421-013-0343-4

Chen, E. H., & Bailey, D. H. (2020). Dual-task studies of working memory and arithmetic performance: A meta-analysis. *Journal of Experimental Psychology : Learning , Memory , and Cognition Working Memory*, *47*(2), 220-233. doi: dx.10.1037/xlm0000822

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. doi: 10.1037/xge0000378

Cosgrave, J., Haines, R., Golodetz, S., Claridge, G., Wulff, K., & van Heugten-van der Kloet, D. (2018). Schizotypy and performance on an insight problem-solving task: The contribution of persecutory ideation. *Frontiers in Psychology*, *9*(MAY), 1–9. doi: 10.3389/fpsyg.2018.00708

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. doi: 10.7771/1932-6246.1129

Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring load on working memory: The use of heart rate as a means of measuring chemistry students cognitive load. *Journal of Chemical Education*, *91*(5), 641–647. doi: 10.1021/ed400576n

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). Working wonders? Investigating insight with magic tricks. *Cognition*, *130*(2), 174–185. doi: 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*(January), 1–14. doi: 10.3389/fpsyg.2016.02077

Danek, A. H., Wiley, J., & Öllinger, M. (2016). Solving classical insight problems without aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *The Journal of Problem Solving*, *9*, 47-57. doi: 10.7771/1932-6246.1183.

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. doi: 10.1037/xlm0000152

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin Books.

Denson, T. F., Grisham, J. R., & Moulds, M. L. (2011). Cognitive reappraisal increases heart rate variability in response to an anger provocation. *Motivation and Emotion*, *35*(1), 14–22. doi: 10.1007/s11031-011-9201-5

Ellis, D. M., & Brewer, G. A. (2018). Aiding the search: Examining individual differences in multiply-constrained problem solving. *Consciousness and Cognition*, *62*, 21–33. doi: 10.1016/j.concog.2018.04.008

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. doi: 10.1037/1082-989X.12.2.121

Estévez-Báez, M., Carricarte-Naranjo, C., Jas-García, J. D., Rodríguez-Ríos, E., Machado, C., Montes-Brown, J., Leisman, G., Schiavi, A., Machado-García, A., Luaces, C. S., & Pié, E. A. (2019). Influence of Heart Rate, Age, and Gender on Heart Rate Variability in Adolescents and Young Adults. In M. Pokorski (Ed.), *Advances in Medicine and Medical Research* (pp. 19–33). Springer International Publishing. doi: 10.1007/5584_2018_292

Fedor, A., Szathmary, E., & Öllinger, M. (2015). Problem solving stages in the five square problem. *Frontiers in Psychology*, *6*(August), 1050–1050. doi: 10.3389/fpsyg.2015.01050

Fedor, A., Zachar, I., Szilágyi, A., & Öllinger, M. (2017). Cognitive architecture with evolutionary dynamics solves insight problem. *Frontiers in Psychology*, *8*(March), 427–427. doi: 10.3389/fpsyg.2017.00427

Fleur, D. S., Bredeweg, B., & van den Bos, W. (2021). Metacognition: ideas and insights from neuro- and educational sciences. *Nature Partner Journals Science of Learning*, *6*(13), 1–11. doi.org/10.1038/s41539-021-00089-5

Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience*, *13*(JUL), 1–11. doi: 10.3389/fnins.2019.00710

Frith, C. D. (2021). The neural basis of consciousness. *Psychological Medicine*, *51*(4), 550–562. doi.org/10.1017/S0033291719002204

Funahashi, S. (2017). Working memory in the prefrontal cortex. *Brain Sciences*, *7*(5), 49. doi: 10.3390/brainsci7050049

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. doi: 10.1037/0033-2909.118.3.392

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. doi: 10.1080/13546780903178615

Gonthier, C., & Roulin, J. L. (2020). Intraindividual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition. *Journal of Experimental Psychology: General*, *149*(3), 564-579. doi.org/10.1037/xge0000660

Hansen, A. L., Johnsen, B. H., & Thayer, J. F. (2003). Vagal influence on working memory and attention. *International Journal of Psychophysiology*, *48*(3), 263–274. doi: 10.1016/S0167-8760(03)00073-4

Harrison, G. M., & Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition Learning*, *13*(April), 15–38. doi.org/10.1007/s11409-017-9176-z

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*(August), 1–13. doi: 10.3389/fpsyg.2016.01314

Hélie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, *117*(3), 994-1024. doi: 10.1037/a0019532

Horvitz, L. A. (2002). *Eureka!: Scientific Breakthroughs that Changed the World*. John Wiley & Sons.

Ingjaldsson, J. T., Laberg, J. C., & Thayer, J. F. (2003). Reduced heart rate variability in chronic alcohol abuse: Relationship with negative mood, chronic thought suppression, and compulsive drinking. *Biological Psychiatry*, *54*(12), 1427–1436. doi: 10.1016/S0006-3223(02)01926-1

Jausovec, N., & Bakracevic, K. (1995). What can heart rate tell us about the creative process? *Creativity Research Journal*, *8*(1), 11–24. doi: 10.1207/s15326934crj0801

Jennings, R. J., Allen, B., Gianaros, P. J., Thayer, J. F., & Manuck, S. B. (2015). Focusing neurovisceral integration: Cognition, heart rate variability, and cerebral blood flow. *Psychophysiology*, *52*(2), 214–224. doi: 10.1111/psyp.12319

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*(4), 500–510. doi: 10.1371/journal.pbio.0020097

Kajić, I., Gosmann, J., Stewart, T. C., Wennekers, T., & Eliasmith, C. (2017). A spiking neuron model of word associations for the remote associates test. *Frontiers in Psychology, 8,* 99-99. doi: www.frontiersin.org/articles/10.3389/fpsyg.2017.00099

Kimhy, D., Crowley, O. V., McKinley, P. S., Burg, M. M., Lachman, M. E., Tun, P. A., Ryff, C. D., Seeman, T. E., & Sloan, R. P. (2013). The association of cardiac vagal control and executive functioning—Findings from the MIDUS study. *Journal of Psychiatric Research*, *47*(5), 628–635. doi: 10.1016/j.jpsychires.2013.01.018

Kounios, J., & Jung-Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*(1), 71–87. doi: 10.1146/annurev-psych-010213-115154

Kounios, John, & Beeman, M. (2015). *The Eureka Factor: Aha Moments, Creative Insights, and the Brain*. Random House.

Kuipers, M., Richter, M., Scheepers, D., Immink, M. A., Sjak-Shie, E., & van Steenbergen, H. (2017). How effortful is cognitive control? Insights from a novel method measuring single-trial evoked beta-adrenergic cardiac reactivity. *International Journal of Psychophysiology*, *119*, 87–92. doi: 10.1016/j.ijpsycho.2016.10.007

Kumral, D., Schaare, H. L., Beyer, F., Reinelt, J., Uhlig, M., Liem, F., Lampe, L., Babayan, A., Reiter, A., Erbey, M., Roebbig, J., Loeffler, M., Schroeter, M. L., Husser, D., Witte, A. V., Villringer, A., & Gaebler, M. (2019). The age-dependent relationship between resting heart rate variability and functional brain connectivity. *NeuroImage*, *185*, 521–533. doi: 10.1016/j.neuroimage.2018.10.027

Laborde, S., Mosley, E., & Mertgen, A. (2018). A unifying conceptual framework of factors associated to cardiac vagal control. *Heliyon*, *4*(12), e01002-e01002. doi: 10.1016/j.heliyon.2018.e01002

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research—Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, *8*(February), 1–18. doi: 10.3389/fpsyg.2017.00213

Lackner, H. K., Weiss, E. M., Schulter, G., Hinghofer-Szalkay, H., Samson, A. C., & Papousek, I. (2013). I got it! Transient cardiovascular response to the perception of humor. *Biological Psychology*, *93*(1), 33–40. doi: 10.1016/j.biopsycho.2013.01.014

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*, *35*(5), 918–935. doi: 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*, 1–6. doi: 10.1016/j.cognition.2019.104122

Laukkonen, R. E., & Tangen, J. M. (2018). How to detect insight moments in problem solving experiments. *Frontiers in Psychology*, *9*(March), 1–5. doi: 10.3389/fpsyg.2018.00282

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. doi: 10.48550/arXiv.1405.2094

Li, Z., Snieder, H., Su, S., Ding, X., Thayer, J. F., Treiber, F. A., & Wang, X. (2009). A longitudinal study in youth of heart rate variability at rest and in response to stress. *International Journal of Psychophysiology*, *73*(3), 212–217. doi: 10.1016/j.ijpsycho.2009.03.002

Lipponen, J. A., & Tarvainen, M. P. (2019). A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *Journal of Medical Engineering and Technology*, *43*(3), 173–181. doi: 10.1080/03091902.2019.1640306

Liu, C., Lin, Y., Ye, C., Yang, J., & He, W. (2023). Alpha ERS-ERD pattern during divergent and convergent thinking depends on individual differences on metacontrol. *Journal of Intelligence*, *11*(4), 1-15. doi.org/10.3390/jintelligence11040074

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 176-201. doi: 10.1037//0278-7393.27.1.176

Magnon, V., Vallet, G. T., Benson, A., Mermillod, M., Chausse, P., Lacroix, A., Bouillon-Minois, J., & Dutheil, F. (2022). Does heart rate variability predict better executive functioning? A systematic review and meta-analysis. *Cortex*, *155*(October), 218–236. doi: 10.1016/j.cortex.2022.07.00

Makovac, E., Thayer, J. F., & Ottaviani, C. (2017). A meta-analysis of non-invasive brain stimulation and autonomic functioning: Implications for brain-heart pathways to cardiovascular disease. *Neuroscience and Biobehavioral Reviews*, *74*(Pt B), 330–341. doi: 10.1016/j.neubiorev.2016.05.001

Marron, R. T., Lerner, Y., Berant, E., Kinreich, S., Shapira-Lichter, I., Hendler, T., & Faust, M. (2018). Chain free association, creativity, and the default mode network. *Neuropsychologia*, *118*(March), 40–58. doi.org/10.1016/j.neuropsychologia.2018.03.018

Martín-Signes, M., Cano-Melle, C., & Chica, A. B. (2021). Fronto-parietal networks underlie the interaction between executive control and conscious perception: Evidence from TMS and DWI. *Cortex*, *134*, 1–15. doi.org/10.1016/j.cortex.2020.09.027

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., & Bates, D. M. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*(June), 305–315. doi.org/10.1016/j.jml.2017.01.001

Meessen, J., Sütterlin, S., Gauggel, S., & Forkmann, T. (2018). Learning by hear—the relationship between resting vagal tone and metacognitive judgments: A pilot study. *Cognitive Processing*, *19*(4), 557–561. doi:10.1007/s10339-018-0865-6

Mednick, S. a. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220–232. doi: 10.1037/h0048850

Merseal, H. M., Luchini, S., Kennet, Y. N., Knudsen, K., Bilder, R. M., & Beaty, R. E. (2023). Free association ability distinguishes highly creative artists from scientists: Findings from the Big-C project. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000545

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. doi: 10.1016/j.tics.2011.07.011

Munoz, M. L., Van Roon, A., Riese, H., Thio, C., Oostenbroek, E., Westrik, I., De Geus, E. J. C., Gansevoort, R., Lefrandt, J., Nolte, I. M., & Snieder, H. (2015). Validity of (ultra-)short recordings for heart rate variability measurements. *PLoS ONE*, *10*(9), e0138921–e0138921. doi: 10.1371/journal.pone.0138921

Nikolin, S., Boonstra, T. W., Loo, C. K., & Martin, D. (2017). Combined effect of prefrontal transcranial direct current stimulation and a working memory task on heart rate variability. *PLoS ONE*, *12*(8), e0181833–e0181833. doi: 10.1371/journal.pone.0181833

Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different?: Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, *39*(1), 36–45. doi: 10.1016/j.intell.2010.11.002

Oberauer, K. (2019). Working memory capacity limits memory for bindings. *Journal of Cognition*, *2*(1), 40-40. doi: 10.5334/joc.86

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity?. *Psychological bulletin*, *142*(7), 758. doi: 10.1037/bul0000046

Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge University Press.

Ohlsson, S. (2018). The dialectic between routine and creative cognition. In F. Vallée-Tourangeau (Ed.), *Insight: On the Orgins of New Ideas* (pp. 8–27). Routledge.

Ottaviani, C., Zingaretti, P., Petta, A. M., Antonucci, G., Thayer, J. F., & Spitoni, G. F. (2018). Resting heart rate variability predicts inhibitory control above and beyond impulsivity. *Journal of Psychophysiology*, *33*(3), 198-206. doi: 10.1027/0269-8803/a000222

Overbeek, T. J. M., Van Boxtel, A., & Westerink, J. H. D. M. (2014). Respiratory sinus arrhythmia responses to cognitive tasks: Effects of task factors and RSA indices. *Biological Psychology*, *99*(1), 1–14. doi: 10.1016/j.biopsycho.2014.02.006

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi: 10.3758/s13428-018-01193-y

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455–473. doi: 10.1037/edu0000079

Quintana, D. S., & Heathers, J. A. J. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, *5*(July), 1–10. doi: 10.3389/fpsyg.2014.00805

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/.

Riddle, J., Scimeca, J. M., Cellier, D., Dhanani, S., & D'Esposito, M. (2020). Causal evidence for a role of theta and alpha oscillations in the control of working memory. *Current Biology*, *30*(9), 1748-1754. doi: 10.1016/j.cub.2020.02.065

Robison, M. K., Miller, A. L., & Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition*, *198*(February), 104078. doi.org/10.1016/j.cognition.2019.104078

Rominger, C., Papousek, I., Fink, A., Perchtold, C. M., Lackner, H. K., Weiss, E. M., & Schwerdtfeger, A. R. (2019). Creative challenge: Regular exercising moderates the association between task-related heart rate variability changes and individual differences in originality. *PLoS ONE*, *14*(7), 1–15. doi: 10.1371/journal.pone.0220205

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. doi: 10.1080/13546783.2016.1141798

Salvi, C., Simoncini, C., Grafman, J., & Beeman, M. (2020). Oculometric signature of switch into awareness? Pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis. *NeuroImage*, *217*(September), 116933-116933. doi: 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), 104038-104038. doi: 10.1016/j.jml.2019.104038

Segerstrom, S. C., & Nes, L. S. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological science*, *18*(3), 275–281. doi: 10.1111/j.1467-9280.2007.01888.x

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(September), 1–17. doi: 10.3389/fpubh.2017.00258

Shen, W., Yuan, Y., Tang, C., Shi, C., Liu, C., Luo, J., & Zhang, X. (2017). In search of somatic precursors of spontaneous insight. *Journal of Psychophysiology*, *32*(3), 97-105. doi: 10.1027/0269-8803/a000188

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. doi: 10.1177/1745691616650647

Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., & Kwapil, T. R. (2014). Creative motivation: Creative achievement predicts cardiac autonomic markers of effort during divergent thinking. *Biological Psychology*, *102*(1), 30–37. doi: 10.1016/j.biopsycho.2014.07.010

Singh, N., Aggarwal, Y., & Sinha, R. K. (2019). Heart rate variability analysis under varied task difficulties in mental arithmetic performance. *Health and Technology*, *9*(3), 343–353. doi: 10.1007/s12553-018-0272-0

Sklar, A. Y., Kardosh, R., & Hassin, R. R. (2021). From non-conscious processing to conscious events: a minimalist approach. *Neuroscience of Consciousness*, *2021*(2). doi.org/10.1093/nc/niab026

Sloan, R. P., Korten, J. B., & Myers, M. M. (1991). Components of heart rate reactivity during mental arithmetic with and without speaking. *Physiology & Behavior*, *50*(5), 1039–1045. doi: 10.1016/0031-9384(91)90434-P

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience and Biobehavioral Reviews*, *75*, 274–296. doi: 10.1016/j.neubiorev.2017.02.003

Smith, T. W., Deits-Lebehn, C., Williams, P. G., Baucom, B. R. W., & Uchino, B. N. (2020). Toward a social psychophysiology of vagally mediated heart rate variability: Concepts and methods in self-regulation, emotion, and interpersonal processes. *Social and Personality Psychology Compass*, *14*(3). doi: 10.1111/spc3.12516

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. doi: 10.5334/irsp.90

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*, 103055-103055. doi: 10.1016/j.concog.2020.103055

Stuyck, H., Cleeremans, A., & Van den Bussche, E. (2022). Aha! under pressure: The Aha! experience is not constrained by cognitive load. *Cognition*, *219*, 104946-104946. doi: 10.1016/j.cognition.2021.104946

Stuyck, H., Dalla Costa, L., Cleeremans, A., & Van den Bussche, E. (2022). Validity of the Empatica E4 wristband to estimate resting-state heart rate variability in a lab-based context. *International Journal of Psychophysiology*, *182*(December), 105–118. doi.org/10.1016/j.ijpsycho.2022.10.003

Tarvainen, M. P., Lipponen, J., Niskanen, J. P., & Ranta-aho, P. O. (2020). Kubios HRV User Guide, Kubios Oy.

Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. doi: 10.1016/j.cmpb.2013.07.024

Tarvainen, M. P., Ranta-aho, P. O., & Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, *49*(2), 172–175. doi: 10.1109/10.979357

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, *37*(2), 141–153. https://doi.org/10.1007/s12160-009-9101-z

Trübutschek, D., Marti, S., Ueberschär, H., & Dehaene, S. (2019). Probing the limits of activity-silent non-conscious working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(28), 14358–14367. doi.org/10.1073/pnas.1820730116

Tukey, J.W. (1977). *Exploratory data analysis*. Addison-Wesley.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. doi.org/10.1016/j.cogpsych.2014.01.003

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(April), 653–673. doi.org/10.3758/s13428-016-0721-5

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). 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. doi: 10.3102/1076998610396895

Webb, M. E., Cropper, S. J., & Little, D. R. (2019). "Aha!" is stronger when preceded by a "huh?": Presentation of a solution affects ratings of aha experience conditional on accuracy. *Thinking and Reasoning*, *25*(3), 324–364. doi: 10.1080/13546783.2018.1523807

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*(September), 1–13. doi: 10.3389/fpsyg.2016.01424

Weisberg, D. S., Kittredge, A. K., Hirsh-Pasek, K., Golinkoff, R. M., & Klahr, D. (2015). Making play work for education. *Phi Delta Kappan*, *96*(8), 8–13. doi: 10.1177/0031721715583955

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, *21*(1), 5–39. doi: 10.1080/13546783.2014.886625

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

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. doi: 10.1080/17470210500234026

Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving In *Psychology of Learning and Motivation* (Vol. 56, pp. 185–227). Elsevier Science & Technology. doi: 10.1016/B978-0-12-394393-4.00006-6

Wilson, D. B. (2022). The relative incident rate ratio effect size for count-based impact evaluations: When an odds ratio is not an odds ratio. *Journal of Quantitative Criminology*, *38*(2), 323–341. doi: 10.1007/s10940-021-09494-w

Wu, C. L., & Chen, H. C. (2021). Distinguishing the three versions of the Chinese Remote Associates Test based on default mode network connectivity. *Thinking Skills and Creativity*, *40*, 100829. doi.org/10.1016/j.tsc.2021.100829

Yoon, J. H., Grandelis, A., & Maddock, R. J. (2016). Dorsolateral prefrontal cortex GABA concentration in humans predicts working memory load processing capacity. *Journal of Neuroscience, 36*(46), 11788–11794. doi: 10.1523/JNEUROSCI.1970-16.2016

Zabelina, D., Saporta, A., & Beeman, M. (2016). Flexible or leaky attention in creative people? Distinct patterns of attention for different types of creative thinking. *Memory and Cognition*, *44*(3), 488–498. doi: 10.3758/s13421-015-0569-4

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

Zeng, J., Meng, J., Wang, C., Leng, W., Zhong, X., Gong, A., Bo, S., & Jiang, C. (2023). High vagally mediated resting-state heart rate variability is associated with superior working memory function. *Frontiers in Neuroscience*, *17*(February), 1–9. doi:10.3389/fnins.2023.1119405

Zhang, W., Sjoerds, Z., & Hommel, B. (2020). Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage*, *210*, 116572-116572. doi: 10.1016/j.neuroimage.2020.116572