Memory & Cognition (2024) 52:115–131
https://doi.org/10.3758/s13421-023-01448-w

Effect of cognitive load and working memory capacity on the efficiency of discovering better alternatives: A survival analysis

Yuki Ninomiya[1] · Tomoyuki Iwata[2] · Hitoshi Terai[3] · Kazuhisa Miwa[2]

Accepted: 14 July 2023 / Published online: 9 August 2023

Abstract

Why do humans attempt to discover better alternatives to solve a problem even when they know the way to solve it? This question is related to the flexibility of thinking and is the subject of studies on the Einstellung effect. This study focuses on cognitive load for familiar trained procedure as a factor that influences the discovery of alternatives in Einstellung situations, where the problem can be solved using a trained procedure. Many studies on creative problem solving and strategy selection demonstrate that cognitive load for facilitates the discovery of alternatives. However, findings are inconsistent regarding Einstellung situations. We argue that the reason for such inconsistent results were the use of manipulations such as the double task, which both load on the trained and alternative procedure and the lack of perspective on efficiency in analysis. Therefore, we examine the relationship between cognitive load manipulated by the complexity of the trained procedure and time-based performance measured by the number of trials prior to finding alternatives. The results illustrate that higher cognitive load increased the efficiency of discovery of alternatives and that this effect was significant for participants with high working memory capacity. This study provides empirical evidence that high cognitive load for trained procedure facilitates the discovery of alternatives and suggests the importance of considering temporal properties, such as efficiency, when examining the effects of working memory on problem solving, which requires cognitive flexibility.

Keywords Einstellung effect · Cognitive load · Working memory · Survival analysis · Problem solving

Humans are capable of flexibly discovering novel ideas and solutions to problems they have never previously experienced apart from known methods (Ohlsson, 2018). This ability, which is called cognitive flexibility (Braem & Egner, 2018; Cañas et al., 2003), is regarded as the capacity to use new methods, techniques, knowledge, and information or to address problems that require the modification of known strategies. It attracts widespread interest in psychological research (Barbey et al., 2013; Bilalić et al., 2008a, b, 2010; Neroni et al., 2017; Sahai & Frese, 2019; Schultz & Searleman, 2002).

However, the discovery of novel ideas and solutions is known to be prevented by a fixation on knowledge and experience (Bilalić et al., 2008a, b, 2010; Haager et al., 2014; Huang et al., 2019; Luchins, 1942; Luchins & Luchins, 1950; Schultz & Searleman, 2002). For example, it is known that when designing creative products, prior knowledge and experience can interfere with the generation of new designs (Jansson & Smith, 1991; Neroni et al., 2017; Okada & Ishibashi, 2017). This phenomenon is called design fixation, and omitting this fixation is considered to lead to the generation of better design. It is also known that fixation on

Yuki Ninomiya, Tomoyuki Iwata, Hitnoshi Terai and Kazuhisa Miwa, have contributed equally to this work.

✉ Yuki Ninomiya
ninomiya.yuki.t1@f.mail.nagoya-u.ac.jp

Tomoyuki Iwata
vdc123tykob315dw@gmail.com

Hitoshi Terai
teraihitoshi@gmail.com

Kazuhisa Miwa
miwa@is.nagoya-u.ac.jp

1 Global Research Institute for Mobility in Society Institutes of Innovation for Future Society, Nagiya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

2 Graduate School of Informatics, Nagiya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

3 Department of Faculty of Humanity-Oriented Science and Engineering, Kindai University, Kayanomori, Iizuka-shi Fukuoka 820-0048, Japan

knowledge and experience leads to decision bias in medical situations (Croskerry, 2003).

The effect of prior knowledge and experience on the prevention of problem solving is called the Einstellung effect (Bilalić et al., 2008a, b, 2010; Luchins, 1942; Luchins & Luchins, 1950). It is a phenomenon in which prior experience or knowledge of a solution are immediately recalled, which, thereby, inhibits the discovery of alternative solutions.

The water jar task (Luchins, 1942; Luchins & Luchins, 1950) is one of the representative tasks related to the Einstellung effect. The objective of this task is to draw a target amount of water using three water jars, —namely, A, B, and C. The task consists of three phases: set, critical, and inspection. In the first set phase, a task that can be solved only with a specific procedure (e.g., B–A–2C) is presented. As a result, this procedure is learned (hereafter, trained procedures). In the following critical phase, the problem can be solved using two procedures: a trained procedure and a simple procedure such as C–A (hereafter, alternative procedures, that is, those that are more efficient or simpler than trained procedures). In the critical phase, discovering simple alternative procedures is difficult, because the trained procedure that immediately comes to mind is applied (Luchins, 1942; Luchins & Luchins, 1950). In this phase, the required amount of water can also be obtained using trained procedures. Therefore, no feedback occurs that certain problems cannot be solved using trained procedures. Lastly, the iInspection phase is one in which the problem can be solved only using alternative instead of trained procedures. Therefore, participants are given a feedback that trained procedures are inappropriate for solving the problem, and they are forced to change the procedure. However, even during the iInspection phase, discovery of alternative procedures is suppressed compared with when trained procedures are not given (Luchins, 1942; Luchins & Luchins, 1950).

Various domains and contexts have confirmed the Einstellung effect. For example, studies have demonstrated this effect using anagrams and chunk decomposition tasks that employ Chinese characters (Ellis & Reingold, 2014; Huang et al., 2019). In addition, functional fixation (Duncker, 1945) and design fixation (Jansson & Smith, 1991; Neroni et al., 2017) have addressed the phenomenon in which knowledge and experience prevent the discovery of new solutions not only in laboratory tasks but also in realistic situations. Previous studies have observed the Einstellung effect in various domains, including studies that focus on the domain knowledge of chess experts (Bilalić et al., 2008a, b, 2010; Sheridan & Reingold, 2013), the use of magic tricks (Thomas et al., 2015; Thomas et al., 2018), and problem solving in mathematics (Chesney et al., 2013). Moreover, scholars have discussed the tendency to continually apply a mental or behavioral set, such as trained procedures, in relation to personality factors such as rigidity (Schultz & Searleman, 2002).

Cognitive load on the discovery of alternative procedures

In this study, we focus on the load on working memory (hereafter referred to as cognitive load) as a factor that facilitates the discovery of alternative procedures. Working memory is a temporary storage system under attentional control and underlies complex thinking such as maintaining, manipulating, and retrieving task-relevant information (Baddeley, 2007; Engle & Kane, 2004; Unsworth & Engle, 2007; Unsworth & Robison, 2017). In analytic problem solving, problem solvers are required to monitor the progress of a task while maintaining mental representations about the problem and avoiding distraction by irrelevant information (Shipstead et al., 2016; Wiley & Jarosz, 2012). Cognitive load reduces the cognitive resources allocated to a task and causes distraction (Eysenck & Calvo, 1992; Eysenck & Derakshan, 2011). Therefore, cognitive load inhibits analytical problem solving by limiting the focus and maintenance of attention on the mental representations of the problem (Gilhooly & Fioratou, 2009; Wiley & Jarosz, 2012).

In contrast, several studies demonstrate that cognitive load facilitates problem solving when easily focalized cues would inhibit problem solving (Ball et al., 2015; DeCaro, 2018; De Caro et al., 2016; Ricks et al., 2007). For example, it is known that in insight problem solving, where relaxing and reconstructing constraints formed by knowledge and experience are necessary steps for finding a solution, high cognitive load and low working memory capacity (WMC) are associated with better performance (Ball et al., 2015; De Caro et al., 2016; Ricks et al., 2007). Ball et al. (2015) illustrated that insight problem solving was facilitated compared with the control condition under conditions of high cognitive load with dual tasks such as articulatory suppression and irrelevant speech. In other words, that the limitation in working memory due to dual tasks inhibits the focus of attention on incorrect information and facilitates problem solving. In addition, DeCaro et al. (2016) report that high WMC (HWMC) are detrimental to insight problem solving, because they retain attention to constraints and inhibits processes specific to insight problem solving such as the reconstruction of problem representations. Ricks et al. (2007) use the remote associates task (RAT) and illustrate a negative correlation between RAT performance and WMC when participants possess knowledge that inhibits problem solving related to the RAT. These findings suggest that problem solvers with ample working memory may excessively focus their attention on incorrect information.

This viewpoint regarding suppression of focusing attention may explain the promotion of the discovery of alternatives even the Einstellung situations. The Einstellung effect is a phenomenon that occurs due to continued attention to trained procedures (Bilalić et al., 2008a, b, 2010). For example, Bilalićc et al. (Bilalić et al., 2008a, b) propose that participants unable to find better alternative procedures focus their attention on known suboptimal procedures recalled from habit (corresponds to trained procedures). The authors then demonstrate that these participants continued to focus their attention on areas related to the suboptimal procedure even when they reported looking for alternatives. In other words, if cognitive load suppresses the focus on trained procedures, then the discovery of alternative procedures may be promoted.

In addition to focusing attention, another perspective explains why cognitive load facilitates the discovery of alternative procedures. In the context of strategy selection, it is known that cognitive load and low WMC(LWMC) encourage the use of shortcut strategies (heuristics), which do not necessarily yield the correct answer, but are less costly than complex strategies that involve detailed computations (Beilock & De Caro, 2007; Caviola et al., 2017; Wang & Shah, 2014; Caviola et al., 2017; Beilock & De Caro, 2007). For example, time pressure, stress, and anxiety facilitate the transition from complex to less costly strategies (Caviola et al., 2017). In addition, a study of arithmetic problems showed that the use of alternative representations was facilitated by the load of the problem (Thevenot & Oakhill, 2005). Specifically, they used arithmetic problems that could be solved using initial representations, which could be efficiently solved by constructing alternative representations. They showed that the use of alternative representations was promoted more for problems requiring three-digit calculations than for those requiring two-digit calculations. These studies suggest that, as the cognitive load of a task increases, a less costly alternative procedure is more likely to be used. In situations, such as the water jar task, the alternative procedure is less costly than the trained procedure. In other words, the discovery of an alternative procedure is likely to be promoted when the cost of using a trained procedure increases owing to the cognitive load.

Influence of cognitive load on the Eeinstellung effect

These previous studies suggest that cognitive load may facilitate the discovery of alternative procedures by suppressing excessive focus on trained procedures. However, previous studies do not reach a conclusion in the discussion of the effect of cognitive load on the discovery of alternative procedures in Einstellung situations (Beilock & DeCaro, 2007; Carr & Steele, 2009; Luchins, 1942). For example, Beilock and DeCaro (2007) found that the proportion of responses under social pressure using alternative procedures was significantly higher in the pressure condition than in the control one condition for participants with HWMC but not significantly different between them for participants with LWMC. Social pressure is known to limit working memory (Beilock et al., 2004; Caviola et al., 2017). In other words, this finding indicates that cognitive load facilitates the discovery of alternative procedures despite being limited to cases with HWMC.

Conversely, under stereotype threat situations, Carr and Steele (2009) illustrated that fixation on trained procedures is encouraged. Similar to social pressure, it is known that working memory is loaded under stereotype threat (Beilock et al., 2007; Inzlicht et al., 2006; Jamieson & Harkins, 2007; Schmader & Johns, 2003). This finding indicates that cognitive load inhibits the use of alternative procedures. In addition, Luchins (1942) reported that more participants responded to trained procedures in the critical phase of the water jar task under the time-pressure condition, in which participants were instructed that their response would be evaluated according to the speed of their response compared with their performance under the control condition. Time pressure was considered one of the common factors that increase cognitive load (Galy et al., 2012). This means that the discovery of alternative procedures was inhibited in situations of cognitive load.

The following points are considered as reasons for the inconclusive results regarding the effect of cognitive load on the discovery of alternative procedures in the Einstellung task:

The first is the problem with the manipulation of cognitive load when examining its influence on the discovery of alternative procedures. Specifically, the manipulation of cognitive load by external pressure or dual tasks makes it impossible to distinguish the load for the use of the trained procedure from that for the search and discovery of an alternative procedure. In Einstellung tasks, such as the water jar task, stop solving the problem using trained procedures and exploring an area unrelated to trained procedures are necessary steps for discovering alternative procedures (Bilalić et al., 2008a, b, 2010; Ninomiya et al., 2022; Sheridan & Reingold, 2013). In the case of a dual task, the exploration of alternative procedures is loaded, even if participants endeavor to explore alternative procedures instead of applying the trained procedures in the water jar task, because the different tasks are performed in parallel with the main task. Therefore, although the cognitive load manipulated by dual tasks encourages the exploration of other procedures, it also increases the difficulty of exploring and discovering alternative procedures, which may not be explicitly manifested as a change in performance. Indeed, it has been shown that the use of alternative procedures is reduced when the cognitive cost of the construction of alternative procedures is high, even for tasks that can be solved more efficiently using alternatives than trained procedures (Thevenot & Oakhill, 2006, 2008). Similar effects have been demonstrated in insight problem-solving tasks, in which initial representations prevent problem solving (Knoblich et al., 2001).

In addition, manipulation by time and social pressure may have influenced the discovery of an alternative procedure as an inhibitory factor, apart from the cognitive load. For example, when people lack confidence in their abilities, they avoid complex strategies under supervision for fear of failure (Laird et al., 2018). In other words, factors other than the cognitive load of evaluation anxiety influenced the choice of problem-solving strategies. Time pressure is also known to influence the estimation of costs and benefits in strategy selection, rather than just the cognitive load (Maule & Hockey, 1993). Specifically, because the existence of an alternative procedure is unknown to the participants in the water jar task, the search for an alternative procedure is a highly uncertain strategy for the goal of finding an answer in as short a time as possible. In other words, time pressure may have affected the goal setting of problem -solving and may have been a disincentive to find alternative procedures.

From the above, it is possible that the manipulation of cognitive load by the double task and social and temporal pressure not only relaxed the focus on the trained procedure, but also inhibited the search and discovery of the alternative procedure. To solve this problem, it is necessary to manipulate the cognitive load on the use of the trained procedure in a manner that does not affect the difficulty in searching for and discovering alternative procedures. However, no previous study has examined the effect of cognitive load on the discovery of alternatives by manipulating only the cognitive load using a trained procedure.

The second pertains to the problem of the lack of perspective on efficiency, such as the time and number of trials required, to discover alternative procedures. Scholars point out the importance of distinguishing between performance effectiveness (e.g., quality of performance and accuracy) and efficiency (e.g., performance effectiveness divided by effort and the amount of effort or time spent on a task) in examining the influence of cognitive load on task performance (Eysenck & Calvo, 1992; Visu-Petra et al., 2013). For example, various studies demonstrate that anxiety, which inhibits its working memory, influences efficiency instead of task effectiveness (Ansari et al., 2008; Visu-Petra et al., 2009). Ball et al. (2015) cite that the effects of thinking aloud and dual tasks on performance in insight problem solving differ between analysis based on solution rates up to an intermediate time point and analysis up to the final time point. These findings indicate that consistent results may not be obtained unless efficiency is considered when examining the influence of cognitive load on problem-solving performance. Therefore, considering the efficiency of discovering alternative procedures is important when examining the influence of cognitive load on the discovery of alternative procedures.

However, previous studies overlooked the influence of cognitive load on the efficiency of discovering alternative procedures (Beilock & De Caro, 2007; Carr & Steele, 2009; Luchins, 1942). For example, Beilock and De Caro (2007) were interested in strategy selection, such as whether or not to use a complex or a simple strategy, and analyzed the frequency of use of alternative procedures during the critical phase as performance data. However, the frequency of use of alternative procedures is not an appropriate indicator of the efficiency of the discovery of alternative procedures due to qualitative differences between the "response using the alternative procedure for the first time" and the "second and subsequent responses using the alternative procedure" in the water jar task. Specifically, using the first alternative procedure requires a departure from fixation on trained procedures to the discovery of the new procedure, whereas using the second and subsequent alternative procedures requires only the application of a known strategy. In addition, Luchins (1942) reported the frequency of whether or not the Einstellung effect was observed, and did not consider efficiency. The discussion of Carr and Steele (2009) was based on the categorical classification of participants who responded with trained procedures at least once in a situation where alternative procedures could be found very easily. Thus, their research interest was not discovering alternative procedures in the first place, and efficiency was also not considered.

Objective

This study aims to examine the influence of cognitive load while using trained procedures on the efficiency of discovering alternative procedures. It defines efficiency as the time variable of how quickly alternative procedures can be discovered. Moreover, we manipulate cognitive load as the complexity of trained procedures instead of dual tasks or pressure.

Experiment 1 examines differences in the efficiency of discovering alternative procedures between high and low cognitive load conditions. It is known that the complexity of a calculation task, which can be manipulated through the inclusion of carries or number of terms, increases cognitive load (Caviola et al., 2012; Noël et al., 2001; Imbo & Le Fevre, 2010; Caviola et al., 2012). In other words, by manipulating the complexity of the calculation of trained procedures, increasing cognitive load is possible when using trained procedures. In contrast, alternative procedures remain the same even when the complexity of trained procedures is manipulated; if cognitive load fort trained procedure encourages the exploration of alternative procedures, then it is expected to lead to an improvement in performance.

To examine the efficiency of discovering alternative procedures, we use the number of trials required to discover alternative procedures as an index. The number of trials until the discovery of alternative procedures correspond to an index of the number of times trained procedures are repeatedly used prior to the discovery of alternative procedures. Therefore, it is considered to be an effective index of efficiency of discovery.

Against this background, the study presents the following hypotheses if cognitive load for trained procedure facilitates the discovery of alternative procedures in the Einstellung task as suggested by the findings of previous studies on attentional focus (Ball et al., 2015; Caviola et al., 2017; DeCaro, 2018; DeCaro et al., 2016; Ricks et al., 2007) and strategy selection (Beilock & DeCaro, 2007; Wang & Shah, 2014; Caviola et al., 2017; Beilock & DeCaro, 2007).

H_1 The more complex the trained procedure, the faster the alternative procedure will be found.

In Experiment 2, we examined the difference in the effects of cognitive load for a trained procedure on the efficiency of discovering alternative procedures depending on whether the WMC is high or low. Previous studies have shown that the effect of cognitive load on strategy selection and problem- solving is dependent on cognitive ability (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Caviola et al., 2017; Wang & Shah, 2014; Caviola et al., 2017; Beilock & DeCaro, 2007). For example, studies have demonstrated that performance in mathematical problem solving is reduced only when participants with HWMC perform high cognitive demand tasks (Beilock & Carr, 2005). Additionally, cognitive load promotes the use of shortcut strategies more strongly in participants with HWMC than in those with LWMC (Caviola et al., 2017; Wang & Shah, 2014; Caviola et al., 2017). Beilock and DeCaro (2007) conducted a water jar task and reported that the facilitating effect of cognitive load on the use of alternative procedures was stronger in participants with HWMC than in those with LWMC.

The effect of cognitive load for trained procedures on facilitating the discovery of alternative procedures may also be more marked for HWMC participants, based on the finding that HWMC is overly focused on known solutions, such as trained procedures (Beilock & De Caro, 2007; DeCaro et al., 2016). However, it is unclear whether there are similar differences in the effects of WMC when only the load on the use of the trained procedure is manipulated, as these WMC-related studies used dual tasks or external pressure. This is because the cause of such a difference in the effect of WMC may be due not only to the load on the use of the trained procedure, but also to the load on the search for alternative procedures. Cognitive load using a dual task is loaded on both the use of trained procedures and the search for alternative procedures. A previous study showed that increased cognitive costs for the construction of alternative procedures inhibits the use of alternative procedures only in LWMC participants (Thevenot & Oakhill, 2006). Therefore, the cognitive load in the search for an alternative procedure may be a confounding factor when examining whether the effect of load on the use of the trained procedure on the discovery of the alternative procedure differs across WMC. In other words, it is necessary to selectively manipulate the cognitive load for trained procedures under the control of the load in the search for alternative procedures to examine whether the effect of the load on the use of trained procedures on the discovery of alternative procedures differs depending on the WMC.

However, no study has organized the effects of cognitive load from this perspective. Furthermore, no study has examined this effect from the viewpoint of the efficiency of discovering alternative procedures. Accordingly, in Experiment 2, we examined whether differences existed in the effect of cognitive load for trained procedures on the discovery of alternative procedures between HWMC and LWMC. If the effect of cognitive load is stronger for participants with HWMC, as reported in previous studies, the following hypothesis can be formulated:

H_{2a} In participants with HWMC, the complexity of the trained procedure facilitates the efficiency of discovering alternative procedures.

H_{2b} In participants with LWMC, the complexity of the trained procedure does not significantly influence the efficiency of discovering alternative procedures.

Experiment 1

Experiment 1 examines the difference in the efficiency of discovering alternative procedures between high and low cognitive load conditions.

This study extended the water jar task of Luchins (1942) to examine the efficiency of discovering alternative procedures from the viewpoint of the number of trials. Specifically, the total number of trials in the water jar task was extended to use the number of trials required to discover alternative procedures as an index of efficiency.

The study manipulated cognitive load using the computational complexity of trained procedures, such as the presence of computational carries and number of terms (Caviola et al., 2012; Imbo & LeFevre, 2010; Noël et al., 2001; Imbo & Le Fevre, 2010; Caviola et al., 2012). The task setting was to obtain the required amount of water using five water jars, —namely, A, B, C, D, and E. The trained procedure for the low-complexity condition is C–D–E (without carrying), and the trained procedure for the high-complexity condition is C–D–D–E (with carrying). The alternative procedure is C–A (without carrying) for both. For example, in the low-complexity condition, the formula 79 – 24 – 41 was set as an answer using the trained procedure for the problem A = 65, B = 63, C = 79, D = 24, and E = 41 (target = 14). In the high-complexity condition, the formula 73 – 17 – 17 – 18 was set as an answer using the trained procedure for the problem A = 52, B = 28, C = 73, D = 17, and E = 18 (target = 21). The corresponding answer using the alternative procedures were 79 – 65 for the low-complexity condition and 73 – 52 for the high-complexity condition.

If cognitive load for trained procedure promotes the discovery of alternative procedures, then the more complex the calculation of the trained procedure, the faster the alternative procedure will be discovered. Therefore, the following results are predicted.

> Prediction 1: The number of trials required to discover alternative procedures will be smaller for the high-complexity condition compared with the low-complexity condition.

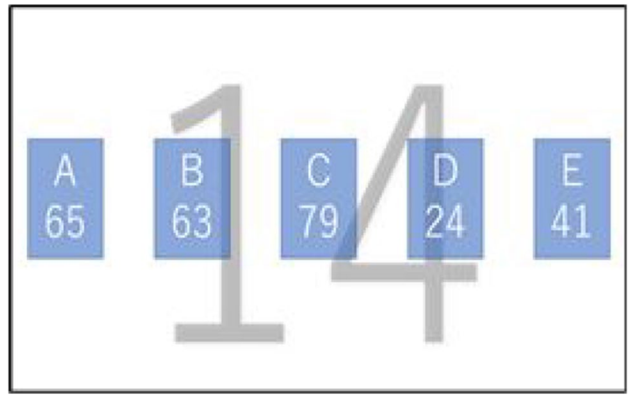

Fig. 1 A sample of the stimulus

Method

Participants

The participants were 74 Japanese university students (N_{female} = 39, N_{male} = 35, M_{age} = 18.36 years, SD_{age} = 0.56 years).

Tasks

The main task was created using jsPsych (de Leeuw, 2015). In the task, five water jars (blue squares) from A to E were displayed, in which the capacity of each water jar was given as a number, and the required amounts of water were provided in the background (Fig. 1). The participants were asked to think of a way to pump the water using the five water jars.

The entire task consisted of 62 trials, which were divided four phases— namely,: induction, set, critical, and inspection. To examine efficiency, the number of critical trials was increased compared to with previous studies because it was necessary to collect early and late finders. The first seven trials were allocated to the induction phase, which encouraged participants to determine the trained procedure. In this phase, each problem could be solved only with the trained procedure, and one of the numbers for D and E or both were divisible by 10 (e.g., in the low-complexity condition C = 71, D = 40, E = 20, and target number = 11). The next eight trials were allocated to the set phase, which could be answered only using the trained procedure. This phase was designed as a learning phase to enable the participants to repeatedly experience that the trained procedure was capable of solving the problem. The difficulty of the trained procedure in the set phase was adjusted to the same difficulty of it in the subsequent critical phase. This setup ensured that the participants would not know the switching from the set to the critical phase given that they solved the problems using the trained procedure. The next critical phase consisted of 45 trials, which could be answered using the trained and alternative procedures. The number of trials from the first trial of the critical phase to the trial answered using the alternative procedure was used as an index of efficiency of discovery. The last two trials (61 and 62) were allocated to the inspection phase, which could be solved using only the alternative procedure. Especially, the final trial was set to C < D, and evidently, the problem could not be solved using the trained procedure. Therefore, final trial in the inspection phase was used as a screening test to test whether or not the participants answered the problems based on their calculations.

Procedure

The experiment was conducted via the Internet. First, the participants received an explanation of the water jar task and given time to practice how to answer it. In this explanation, participants were instructed to answer how to draw the background target amount of water using five water jars. After the practices, the participants conducted the main task. Participants were randomly assigned to the low or high-complexity condition. Finally, after completing the task, the participants were asked whether or not they experienced the water jar task in the past.

Results

Screening of Pparticipants

One participant who had never solved using the trained procedure in the set phase was excluded, because this participant could be considered unable to form a mental set. In the last trial of the inspection phase, the trained procedure could evidently not be applied if the participants solved the task seriously. The four participants who answered this trial using the trained procedure were excluded, because they were considered to have entered their answers without performing any calculations. Finally, 69 participants (N_{female} = 38, N_{male} = 31, M_{age} = 18.33 years, SD_{age} = 0.50 years) were included in the analysis.

Analysis of the efficiency of discovering alternative procedures using survival analysis

To verify Prediction 1, the study employed survival analysis. It is a method of analyzing data in which the time until an event occurs is the target variable (Kleinbaum & Klein, 2012). This analysis is frequently used in fields, such as medicine, to analyze the time spent until the onset of symptoms or death; thus, it is called *survival* analysis. Using this method, the proportion of discovery of alternative procedures within a defined time period can be analyzed as a cumulative proportion and compared across conditions. In addition, censored data for participants who were unable to determine alternative procedures during the last trial can be included in the time-series analysis instead of simply as undiscovered data. For these reasons, this analysis is considered suitable for the objective of the study.

In the analysis, the discovery of the alternative procedure in the critical phase were regarded as the event, and the participants who could not find it until the end of the critical phase (up to 45 trials) were regarded as censored data. Survival time was defined as the number of trials until the occurrence of the event. The effect of the complexity of the calculation on the survival time was then examined. First, we plotted the survivor curves for each condition using the Kaplan–Meier method (Fig. 2). Survivor curves reflected the cumulative proportion of the participants who discovered alternative procedures by the time of each trial. In addition, the hazard functions for each condition were plotted to examine the manner in which the effect of conditions on the number of trials until discovery changed (Fig. 3). The hazard rate (conditional failure rate) reflected the rate of occurrence of the event under the assumption that the event (discovery of alternative procedures in this study) did not occur until a certain point in time.

Wilcoxson's test was then performed to examine the differences in survivor curves in Fig. 2. Analysis revealed a significant difference in survivor curves between complexity conditions, ($\chi^2(1) = 4.97$, $p = .028$). This finding indicated discovering alternative procedures in the high-complexity condition was easier than that in the low-complexity condition.

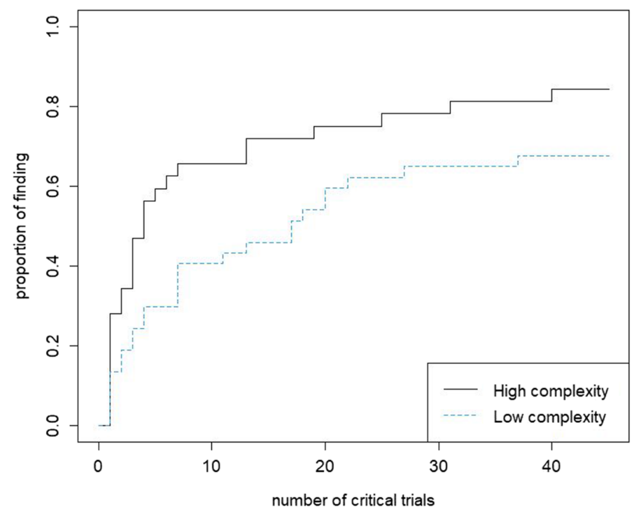

Fig. 2 Comparison of Kaplan–Meier survivor curves between complexity conditions in Experiment 1

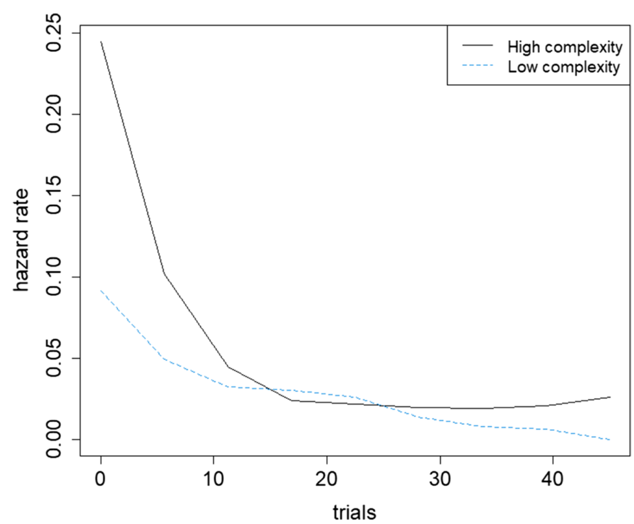

Fig. 3 Comparison of hazard rates between complexity conditions in Experiment 1

Based on the plot of hazard rates in Fig. 3, the study also observed that regardless of complexity condition, the discovery of alternative procedures tended to have high hazard rates in the early stages and a decreasing rate in the latter stages.

Finally, we examine whether or not the hazard ratio changes over time. The hazard ratio is the ratio of hazard rates across conditions. A large hazard ratio at given time t indicates that at time t, the effect of the difference between conditions on the rate of occurrence of the event is large. Therefore, the hazard ratio between complexity conditions point to the magnitude of the effect of the complexity on the discovery of alternative procedures. If the hazard ratio is constant over time, then the effect of complexity is constant over all trials. To examine this point, the study creates a Cox proportional hazards regression model to explain survival time under the complexity condition. Moreover, it conducts a goodness-of-fit (GOF) test using the proportional hazards assumption that "the hazard ratio is constant over time" as the null hypothesis (Grambsch & Therneau, 1994). In this test, if the null hypothesis is rejected ($p < .05$), then the assumption that "the hazard ratio is constant over time" is rejected. Conversely, if the null hypothesis is insufficiently rejected ($p > .10$), then the assumption that the "hazard ratio is constant over time" is considered satisfied (Kleinbaum & Klein, 2012). The results of the analysis indicate no significant results, ($\chi^2(1) = 0.70$, $p = .402$). In other words, the assumption that the hazard ratio is constant over time was not rejected.

Finally, we examined whether there were differences in discovery performance between conditions when efficiency was not considered. Specifically, we compared the proportion of participants who found alternative procedures to with those who did not find alternative procedures in the critical trial across conditions. In the low-complexity condition, 25 participants were found, and in the high-complexity condition, 27 participants were found. Fisher''s exact test result showed no significant differences ($p = .161$, $\varphi = 0.16$).

Summary

The result demonstrated that the number of trials required to find alternative procedures was decreased under the high-complexity condition compared with the low-complexity condition as predicted in Prediction 1. It provides empirical evidence that cognitive load for trained procedure promotes the efficiency of discovery of alternative procedures. Therefore, H_1 is supported.

In addition, the study observed that the hazard rate of discovering alternative procedures decreased with the increase in survival time. This notion implied that the likelihood of discovering alternative procedures is higher during the early trials and lower in the later trials. This result may reflect the temporal nature of the Einstellung effect, —that is, fixation on the trained procedure increases with the progress in the task, although this argument is not based on statistical significance.

Finally, the proportional hazards assumption that the hazard ratio remains constant over time was not rejected. In other words, no evidence exists that the effect of cognitive load on the number of trials required to discover alternative procedures varied over time.

Experiment 2

The results of Experiment 1 provided empirical evidence that cognitive load for trained procedure promotes the efficiency of discovery of alternative procedures. Experiment 2 examined the manner in which the effect of cognitive load for trained procedure on the discovery of alternative procedures differs between HWMC and LWMC with a focus on the efficiency of discovery. Several types of WMC measurement methods are available, which exert varying impacts on problem-solving performance (Van Stockum Jr & De Caro Jr., 2020). Therefore, this study used the Operation Span Test (OSPAN; Kobayashi & Okubo, 2014; Turner & Engle, 1989, Kobayashi & Okubo, 2014). Previous studies demonstrated that OSPAN can interact with cognitive load on performance in the water jar task (Beilock & De Caro, 2007).

Additionally, Experiment 2 intended to verify whether or not the results of Experiment 1 can be replicated in a diversified group of participants using crowdsourcing. The participants in Experiment 1 was limited to university students, and the age variance was very narrow ($SD_{\text{age}} = 0.50$). It is known that age influences the tendency to maintain one's mental set (Schultz & Searleman, 2002). Therefore, a possibility exists that the promoting effect of cognitive load for trained procedure on the discovery of alternative procedures in Experiment 1 is a result specific to the age group. In Experiment 2, we will first verify whether or not the effect of cognitive load for trained procedure on the discovery of alternative procedures is replicable in a diverse group of participants.

After verifying the replication of this effect, the following predictions are considered, which can be made corresponding to H_{2a} and H_{2b} based on the finding that the change in performance by the cognitive load is stronger for participants with HWMC than for those with LWMC (Beilock & DeCaro, 2007; Caviola et al., 2017; Wang & Shah, 2014; Caviola et al., 2017; Beilock & DeCaro, 2007).

> Prediction 2a: In participants with HWMC, the number of trials required to find alternative procedures will be smaller under the high-complexity condition compared with the low-complexity condition.
> Prediction 2b: In participants with LWMC, the complexity of trained procedure will no significantly influence the number of trials required to find alternative procedures

Method

Participants

In Experiment 2, the study employed a Japanese internet crowdsourcing service to recruit 188 participants from diverse backgrounds (N_{female} = 84, N_{male} = 104, M_{age} = 38.64 years, SD_{age} = 8.56 years). Participants could participate in the experiment by submitting an internet form and received a reward for their participation. All participants were Japanese.

Tasks

Main Task

The tasks for Experiment 2 were also created in jsPsych (de Leeuw, 2015). The overall number of trials in this task was reduced to 41. Specifically, it consisted of two trials in induction phase, seven trials in the set phase, 30 trials in the critical phase, and two trials in the iInspection phase. In the first four trials of all trials, after their responses, the participants received feedback on how to solve problems using the trained procedure. The purpose of this feedback is to reduce the number of participants who do not discover the trained procedure during the set phase (for an example of using feedback, see Van Stockum Jr & De Caro Jr., 2020). The rest of the experimental procedure was the same as in that in Experiment 1.

OSPAN

The operation span (OSPAN) test is a task in which computation and word memorization are repeated in parallel, and WMC is measured using word reproduction performance. The study used the Japanese version of OSPAN (Kobayashi & Okubo, 2014), because the experiment was conducted on Japanese participants. This test was not developed to be performed online. However, it has been shown that there is a positive correlation between web and laboratory OSPAN performance (Ruiz et al., 2019). Therefore, it is considered suitable for online use.

The test consisted of four phases, —namely, computation, memorization, reproduction, and feedback. In the computation phase, the participants confirmed whether or not the answer to a formula presented on the screen was the same as or different from a number presented on the next screen. In the subsequent memorization phase, the participants memorized the alphabet displayed for 800 ms. After repeating the computation and memorization for a designated number of spans (3–8), the participants were asked to reproduce the alphabet they memorized in the reproduction phase. The number of spans was unknown to the participants at this point. Finally, after providing responses in the reproduction phase, the answers were feedback in the feedback phase. This sequence of phases was defined as one set and repeated. The OSPAN score was calculated using the partial scoring method (Kobayashi & Okubo, 2014; Unsworth et al., 2005), in which scores are added for each correct reproduction of the presented an alphabet. For example, if “"A, B, C, D”" is presented and participant answers “"A, E, C, B",” 2 points are added because A and C are correct in position. This score was used as the WMC score of the participants.

Procedure

The participants followed the instructions on the cloud service. First, they entered their demographic information (e.g., sex and age), then performed the main task and OSPAN. The procedure for the main task was the same as that in Experiment 1. OSPAN was conducted after the completion of the main task.

Results

Screening of Participants

The participants were screened using the same criteria as those in Experiment 1. Six participants who never responded using the trained procedure in the set phase and 15 participants who responded using the trained procedure in the second trial of the iInspection phase were excluded. A total of 167 participants were included in the analysis (N_{female} = 74, N_{male} = 93, M_{age} = 39.01 years, SD_{age} = 8.42 years).

Replication of the Result of Experiment 1

Survival curves and hazard functions were plotted in the same manner as those for Experiment 1 (Fig. 4 and 5, respectively) to examine whether or not the results of Experiment 1 could be replicated for diverse participants. First, Wilcoxson's test was performed to examine the differences in survival curves between complexity conditions. The analysis revealed a significant difference between the two survival curves, ($\chi^2(1) = 6.43$, p = .011), which indicates that the likelihood of discovering alternative procedures was higher under the high-complexity condition than that in the low-complexity condition, which is similar to Experiment 1. The study then observed a trend in the plots, that is, the hazard rate of alternative procedures was higher in the early stages of the task. Subsequently, the study created the Cox proportional hazards regression model to explain survival time under the complexity conditions and conducted a GOF test using the proportional hazards assumption that “the hazard ratio is constant over time” as the null hypothesis (Grambsch & Therneau, 1994). The results demonstrated that the proportional hazards assumption was not rejected, ($\chi^2(1) = 2.36$, p = .125). Finally, we then compared the proportion of finders to non-finders across conditions to examine whether there were any differences in discovery performance across conditions when efficiency was not considered. In the low-complexity condition, 41 participants were

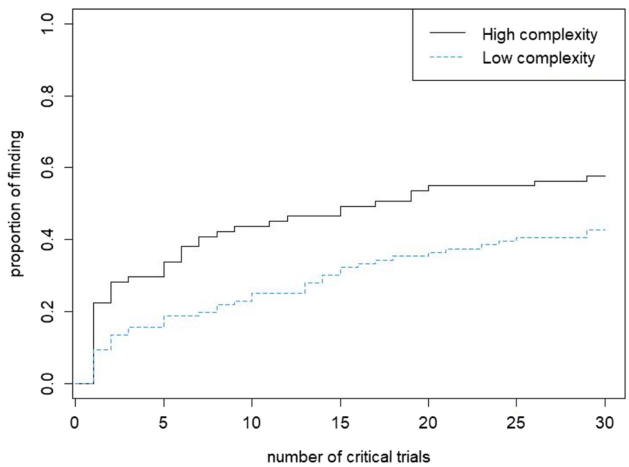

Fig. 4 Comparison of Kaplan–Meier survivor curves between complexity conditions in Experiment 2

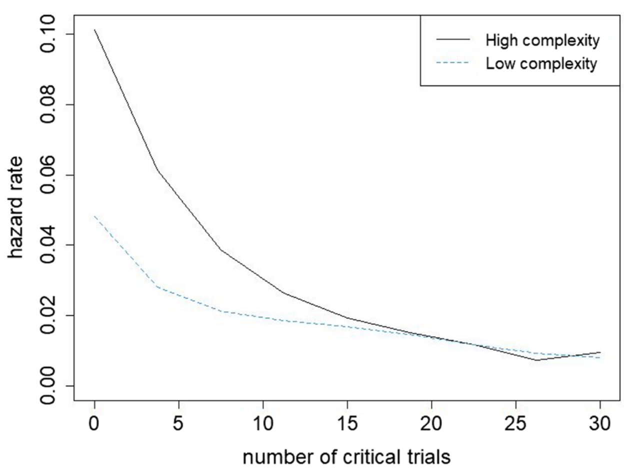

Fig. 5 Comparison of hazard rates between complexity conditions in Experiment 2

found, and in the high-complexity condition, 41 participants were found. Fisher's exact test results showed no significant differences ($p = .062$, $\varphi = 0.14$).

Differences in the effect of complexity on survival time between HWMC and LWMC

The study examined whether or not the effect of complexity on the number of trials required to discover alternative procedures differs between HWMC and LWMC. The participants were classified as the HWMC group ($N_{female} = 35$, $N_{male} = 49$, $M_{age} =$ 37.98 years, $SD_{age} = 8.46$ years) and the LWMC group ($N_{female} = 39$, $N_{male} = 44$, $M_{age} = 40.05$ years, $SD_{age} = 8.31$ years) using the median score (75). Figure 6 presents the survival curves and hazard functions under the high- and low-complexity conditions.

First, the study separately performed the Wilcoxon test for the HWMC and LWMC groups to examine differences in survival curves according to complexity. The results demonstrated that a significant difference exists for participants with HWMC, ($\chi^2(1) = 5.82$, $p = .016$,) but not for those with LWMC, ($\chi^2(1) = 1.69$, $p = .193$). Specifically, for the HWMC group, alternative procedures were more likely to be discovered under the high-complexity condition than the low-complexity condition, whereas no significant difference was observed in the survival curves of the LWMC group between the two conditions.

Subsequently, the study examined whether the hazard ratio (the ratio of hazard rates across conditions) changed over time based on the plotted hazard functions and GOF test. Based on the plots of the hazard rate, the study observed that the hazard rate always tended to be higher under the high-complexity condition for the participants with HWMC. Moreover, the GOF test did not reject the null hypothesis that the hazard rate was constant over time, ($\chi^2(1) = 0.49$, $p = .484$), which is similar to Experiment 1. Alternatively, on the basis of the plots of the hazard rate, the study observed that the relationship between hazard ratios across conditions tended to be reversed in the middle of the task for participants with LWMC. However, the results of the GOF test presented no significant difference, ($\chi^2(1) = 3.77$, $p = .052$).

Summary

First, the study examined the effect of complexity on the efficiency of discovery in the diverse participant group, and the result of Experiment 1 was replicated. Specifically, the effect of the complexity of the trained procedure is independent of specific age groups, which indicates the robustness of the effect.

Subsequently, according to Prediction 2a and 2b, the effect of complexity was observed only for participants with HWMC. These results support H_{2a} and H_{2b} and are consistent with previous findings, that is, the effects of cognitive load on problem-solving performance and strategy selection are stronger for participants with HWMC than those with LWMC (Beilock & DeCaro, 2007; Caviola et al., 2017; Wang & Shah, 2014; Caviola et al., 2017; Beilock & DeCaro, 2007). The important points are that only the cognitive load on the trained procedure was manipulated and that efficiency performance was used. In other words, these results newly confirmed that this trend is observed in the effect of cognitive load for trained procedure on the efficiency of discovering alternative procedures.

General discussion

Effect of cognitive load on the efficiency of discovering alternative procedures

This study examined the effect of the complexity of the trained procedure on the efficiency of discovering alternative

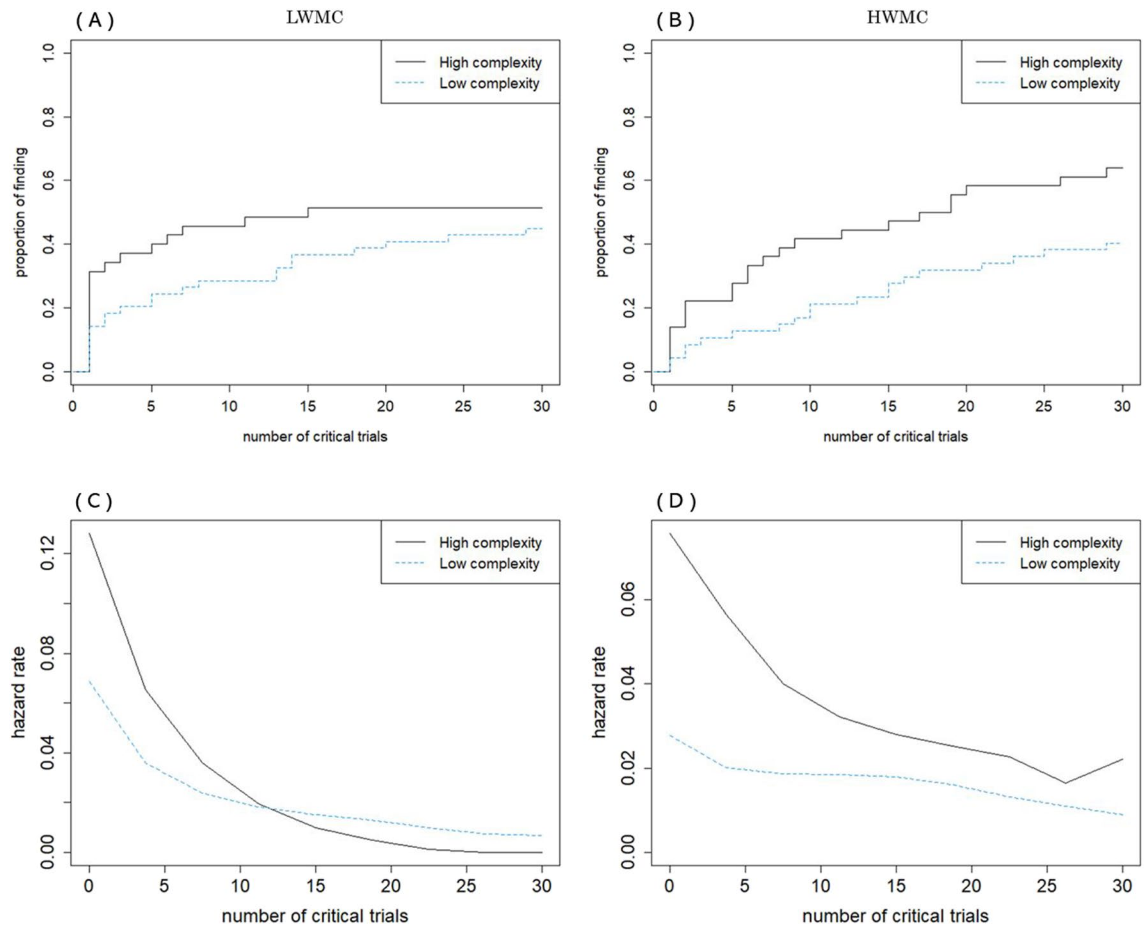

Fig. 6 Kaplan–Meier survival curves and hazard rates by HWMC and LWMC

procedures with a focus on cognitive load for trained procedure as a factor that promotes deviation from the Einstellung effect.

Experiment 1 demonstrated that alternative procedures are more likely to be discovered under the high-complexity condition of the trained procedure than under the low-complexity condition, which is consistent with H1. This tendency was also time independent. These results were replicated in Experiment 2, which was conducted with participants from a wide age range.

These results indicate that alternative procedures are more likely to be discovered when there is a high cognitive load for the trained procedure. This effect was replicated across different age groups, suggesting that the results are robust, and at least independent of age group or participant status (e.g., student or not).

These results provide detailed evidence of the role of cognitive load in terms of the Einstellung effect. Previous studies (Beilock & DeCaro, 2007; Carr & Steele, 2009; Luchins, 1942) may have confounded the effects of load on the trained procedure and the effect of load on the alternative procedure because of the use of dual tasks and external pressure to manipulate cognitive load. Specifically, the dual task and external pressure not only loaded the use of the trained procedure, but might also load the search for an alternative procedure. In contrast, the present study was designed to apply cognitive load only to the use of trained procedures so that the cognitive load would not affect the search for alternatives. Therefore, the results provide evidence that the cognitive load associated with the use of trained procedures (initial solutions), as acquired through learning and experience, facilitates the discovery of alternative procedures.

The results of this study indicate that both the cognitive load on the use of trained procedures and the cognitive load on the search for alternative procedures must be considered to explain the effect of cognitive load on the discovery of alternatives. However, the influence of cognitive load on the search for alternative procedures has not been examined in this study. Previous studies related to insight have shown that the higher costs of constructing alternative representations inhibit their use and discovery of alternative representations (Knoblich et al., 2001; Thevenot & Oakhill, 2006, 2008). Therefore, the cognitive load in the search for alternative procedures may inhibit discovery, in contrast to the cognitive load in the use of trained procedures. The reason that cognitive load inhibited the discovery of alternative procedures in previous studies (Carr & Steele, 2009; Luchins, 1942) may be that the inhibitory effect of the load on the search for alternatives exceeded the facilitatory effect of the load on the trained procedure in the case of manipulation by time pressure or stereotypic threat. In future studies, it will be possible to have a more detailed description of the relationship between cognitive load and the Einstellung effect by examining the effects of manipulating only the load on the search for alternative procedures in the Einstellung task on the discovery of alternative procedures.

In addition, the present results suggest that the complexity of the procedure encourages the search for alternatives even when a well-known procedure can successfully solve the problem. This notion implies that people may evaluate procedures even in situations where a well-known procedure achieves the goal of the problem. In fact, it is known that in problem solving and reasoning, people automatically evaluate their answers and judgments regardless of whether or not they are correct (Ackerman & Thompson, 2017). For example, previous studies have shown that when one's heuristic judgment is low in fluency, that judgment is more likely to be perceived as incorrect and consequently abandoned (Alter et al., 2007; Thompson et al., 2013, 2015). In other words, it suggests that metacognitive evaluation works even for intuitive judgments based on empirically acquired heuristics. The present findings imply that even when people know a procedure that is evidently conducive to solving the problem, such as the trained procedure, they can evaluate the procedure and search for alternatives. In future studies, examining the manner in which people evaluate trained procedures and the influence of such evaluations on the discovery of alternative procedures are necessary aspects.

Relationship between working memory capacity and effect of cognitive load on the efficiency of discovering alternative procedures

The second objective of this study was to examine differences in the effects of cognitive load for trained procedure on the discovery of alternative procedures between participants with HWMC and LWMC. The results of survival analysis in Experiment 2 illustrated differences in the effects of cognitive load between participants with HWMC and LWMC. Specifically, the participants with HWMC were more likely to find alternative procedures when the cognitive load for trained procedure was higher, whereas no difference was noted for participants with LWMC, which supported H_{2a} and H_{2b}.

These results are consistent with previous findings that the promotional effect induced by inhibiting focus in participants with HWMC is stronger than in those with LWMC since participants with HWMC are more likely to focus on complex solutions. (Beilock & DeCaro, 2007; Ricks et al., 2007). From the perspective of attentional focus, the results can be interpreted as follows: Cognitive load inhibited excessive focus on the trained procedure, even in the Einstellung situation (Bilalić et al., 2008a, b, 2010) and, thereby, facilitating the discovery of alternative procedures. In other words, the reason for the different effectiveness of the different WMCs is that HWMC focuses more on the trained procedure than LWMC.

Moreover, this result is consistent with previous studies that show that HWMC is more likely to rely on high-cost strategies (Beilock & DeCaro, 2007; Caviola et al., 2017; Beilock & DeCaro, 2007; Fischer & Holt, 2017; Wang & Shah, 2014). Based on the perspective of strategy selection, the reason underlying the effect of cognitive load for trained procedure may be that the cost of continuing to use the trained procedure increases with the increase in complexity, which leads to the discovery of alternative procedures with a low cost. The reason for the difference in the effect of load on the trained procedure on the discovery of alternatives is explained by the fact that HWMC is more likely to rely on complex procedures before a cognitive load is required, whereas LWMC prefers simple procedures (Beilock & DeCaro, 2007; Caviola et al., 2017; Beilock & DeCaro, 2007; Fischer & Holt, 2017).

Importantly, the cognitive load used in this study was based only on the use of trained procedures. This means that differences in the facilitation effects among WMCs arise in relation to cognitive load during the use of trained procedures. This finding is important when considering the causes of differences in the effects of cognitive load among WMCs. The cognitive load caused by the dual task loads not only the use of trained procedure, but also the search for alternative procedure. Therefore, previous studies have confounded the possibility that differences in the effects of cognitive load among WMCs are caused by differences in the effects of load on the search for alternative procedures. In other words, it is possible that the cause of the differences in effects among the WMCs was the load on the search for alternative procedures, and not the load on the use of trained procedures. Indeed, an increase in the cognitive cost of constructing alternative procedures inhibits the use of alternative procedures only for participants with LWMC (Thevenot & Oakhill, 2006). Based on this finding, even if the effect of load on the use of the trained procedure was the same across WMCs, differences in the effect across WMCs would be expected under the cognitive load of the dual task. However, even in the present study, in which only the use of the trained procedure was loaded, the effect of cognitive load on discovery was pronounced in HWMC. In this sense, it provides direct evidence that the cognitive load on trained procedures is related to the phenomenon in which the effect of cognitive load is more pronounced in HWMC.

This supports the explanation that the effect of cognitive load on performance is more pronounced in HWMC because HWMC participants are more likely to focus on trained procedures and prefer costly and complex strategies. In future studies, it will be necessary to examine the relationship between cognitive load for each procedure and the discovery of alternative procedures, and the interaction between these effects and WMC.

Moreover, the present results can be explained in terms of curiosity. Previous studies illustrate that task complexity reduces exploitation and promotes exploration (Dubey & Griffiths, 2020; Minton, 1963). To discover alternative procedures under the situation where the Einstellung effect is observed, ceasing initial responses using trained procedures based on prior knowledge, and exploring alternatives are necessary steps (Bilalić et al., 2008a, b, 2010; Ninomiya et al., 2022; Sheridan & Reingold, 2013). Therefore, the study infers that exploitation corresponds to the use of trained procedures, and exploration denotes the exploration of alternative procedures. Specifically, the study interprets the result as the high complexity of trained procedures encouraged the exploration instead of the utilization of information, which, in turn, encouraged the discovery of alternative procedures.

This curiosity perspective can also explain the lack of the facilitating effect of cognitive load for the participants with LWMC. It is known that the promoting effect of complexity on information search exhibits an inverse U shape (Baranes et al., 2014; Kidd et al., 2012; Kidd & Hayden, 2015). In other words, the facilitating effect of exploration is not observed under excessively high-complexity conditions. Although the complexity of trained procedures was objectively the same, it may have been subjectively higher for the participants with LWMC compared with those with HWMC. Specifically, the participants with LWMC may perceive the task as excessively complex under the high-complexity condition compared with those with HWMC, which may not facilitate the search for alternative procedures.

In future studies should focus not only on attentional focus and strategy selection but also on changes in the extent of exploration from the viewpoint of curiosity. In particular, another necessary aspect is the investigation of whether or not the interpretation of the lack of the facilitating effect of complexity for participants with LWMC is because "they did not excessively focus on trained procedures and, therefore, attentional distribution exerted less promoting effect" or "the high-complexity condition made the task excessively complicated, such that exploration was not encouraged" by manipulating the degree of cognitive load in a step-by-step manner.

Importance of considerations based on the efficiency of the discovery of alternative procedures

The current results demonstrated that cognitive load for trained procedure influences the discovery of alternative procedures under the Einstellung effect. The important point here is that these results indicate that cognitive load influences the efficiency of discovery by analyzing the number of trials until discovery as survival time.

For example, the results of survival analysis pointed to a difference in the survival curves for the high- and low-cognitive load conditions (Fig. 2, 3, 4, and 5). Furthermore, the study analyzed survival time as the number of required trials until the discovery of alternative procedures. The difference between the two survival curves suggested that more participants found alternative procedures on earlier trials under the high-complexity condition with a high cognitive load than under the low-complexity condition. This finding suggested that cognitive load for trained procedure promotes the efficiency of discovering alternative procedures in the sense that the higher the cognitive load for trained procedure, the smaller the number of trials until the discovery of alternative procedures.

The plots of hazard rates (Fig. 3 and 5) indicate that the likelihood of discovering alternative procedures was larger in the early trials and smaller in the later trials. This aspect suggests the temporal nature of the Einstellung effect, although this argument is not based on statistical significance. Crooks and McNeil (2009) report that the number of trials in the set phase indicates the strength of fixation to trained procedures. Based on these findings, the study considers that this tendency may be attributed to an increase in positive examples for trained procedures as the latter trials progress, and fixation on trained procedures becomes stronger.

For example, when we only analyzed the ratio of finders without considering efficiency, there was no significant difference between the complexity conditions in Experiments 1 and 2. In other words, although the effect of cognitive load was shown for survival time as an efficiency measure, no difference was found for discovery rate as an effectiveness measure. Furthermore, the effect of cognitive load on the discovery of alternative procedures in LWMC was reversed between the early and late stages of the trial, according to the plot of hazard rate (Fig. 6 C). This result suggests that the effect of cognitive load differed by time, although not enough to reject the proportional hazards assumption. Ball et al. (2015) also demonstrated such a change in the effect of cognitive load according to survival time in insight problem solving. Therefore, a possibility exists that the difference in the effects between HWMC and LWMC can be attributed to this change in the effect over time. This discussion of changes in the hazard rate and hazard ratio, although not demonstrated using statistical tests, suggests the importance of considering temporal concepts, such as efficiency, when examining the effects of cognitive load on performance under Einstellung situations.

The discussion of these temporal changes in performance can only be demonstrated by examining efficiency such as the number of required trials until discovery. Several studies that addressed temporal changes in problem-solving processes point out the importance of exploring the temporal features of problem-solving performance (Ball et al., 2015; Sowden et al., 2015). Nevertheless, few reports on the introduction of the efficiency perspective in the process of finding alternatives under the Einstellung effect are available. Therefore, the contribution of the findings to previous studies is the demonstration of the importance of efficiency-based arguments in considering the effect of cognitive load and WMC on the discovery of alternative procedures.

Future work should also examine the effect of cognitive load on the efficiency of the discovery of alternatives in various variations of the Einstellung task. For example, some Einstellung tasks utilize trained procedures that humans have acquired empirically over a longer span of time (Bilalić et al., 2008a, b, 2010; Sheridan & Reingold, 2013). Trained procedures for such tasks may be more automated and require less effort than those created in a shorter period of time, such as the trained procedure for the water jar task. Therefore, such tasks may be less sensitive to the cognitive load on trained procedures. As another example, the instructions given to the participants also depend on the type of Einstellung task. Our task did not require participants to respond by using the smallest number of steps. However, in the traditional water jar task, participants were asked to perform the minimum number of steps (Luchins, 1942). In such cases, the effect of cognitive load may be smaller because participants are motivated to search for alternative procedures regardless of the cognitive load. The examination of such a variety of tasks would allow a detailed description of the effect of cognitive load on the efficiency of discovering alternatives and would help explain why the effect of cognitive load is inconsistent.

Conclusion

This study examined the effect of cognitive load on the efficiency of discovering alternative procedures in a situation wherein a problem can be solved using well-known procedures. The results of the two experiments illustrated that survival time until the discovery of alternative procedures differed according to the complexity of trained procedures and that survival time was shorter under the high-complexity condition. The study confirmed this effect among the participants with HWMC but not the those with LWMC. These results provided empirical evidence that cognitive load for trained procedure on people with HWMC promotes the efficiency of discovering alternative procedures. In conclusion, this paper adds to the findings on the effects of cognitive load and WMC on discovering alternative procedures in terms of temporal efficiency in Einstellung situations.

Authors' contributions All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding This work was supported by JSPS KAKENHI (Grant Number No. 22H03912).

Data availability The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Code availability Not applicable.

Declarations

Conflicts of interest/Competing interests We know of no conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome.

Ethics approval This study was conducted with the approval of the ethics committee of the affiliated institution.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Participants signed informed consent regarding publishing their data. All authors have approved the manuscript and agree with its submission to *Memory and & Cognition.*

Open Practices Statement None of the experiments was preregistered.

Reference

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Mmetacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Ansari, T. L., Derakshan, N., & Richards, A. (2008). Effects of anxiety on task switching: Evidence from the mixed antisaccade task. *Cognitive, Affective, & Behavioral Neuroscience, 8*(3), 229–238. https://doi.org/10.3758/CABN.8.3.229

Baddeley, A. (2007). *Working memory, thought, and action* (45th ed.). OuP Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., & Booth, N. (2015). When distraction helps: Evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Thinking & Reasoning, 21*(1), 76–96. https://doi.org/10.1080/13546783.2014.934399

Baranes, A. F., Oudeyer, P. Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in nNeuroscience, 8*, 317. https://doi.org/10.3389/fnins.2014.00317

Barbey, A. K., Colom, R., & Grafman, J. (2013). Architecture of cognitive flexibility revealed by lesion mapping. *NeuroIimage, 82*, 547–554. https://doi.org/10.1016/j.neuroimage.2013.05.087

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Sscience, 16*(2), 101–105. https://doi.org/10.1111/j.0956-7976.2005.00789.x

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 983–998. https://doi.org/10.1037/0278-7393.33.6.983

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*(4), 584–600. https://doi.org/10.1037/0096-3445.133.4.584

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256–276. https://doi.org/10.1037/0096-3445.136.2.256

Bilalić, M., McLeod, P., & Gobet, F. (2008a). Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Ppsychology, 56*(2), 73–102. https://doi.org/10.1016/j.cogpsych.2007.02.001

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the Einstellung (Set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science, 19*(2), 111–115. https://doi.org/10.1177/0963721410363571

Bilalićc, M., McLeod, P., & Gobet, F. (2008b). Why good thoughts block better ones: The mechanism of pernicious Einstellung (Set) effect. *Cognition, 108*, 652–661. https://doi.org/10.1016/j.cognition.2008.05.005

Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current dDirections in Psychological Science, 27*(6), 470–476. https://doi.org/10.1177/0963721418787475

Cañas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics, 46*(5), 482–501. https://doi.org/10.1080/0014013031000061640

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*(4), 853–859. https://doi.org/10.1016/j.jesp.2009.03.003

Caviola, S., Mammarella, I. C., Cornoldi, C., & Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *Journal of Experimental Child Psychology, 112*(2), 141–160. https://doi.org/10.1016/j.jecp.2012.02.005

Caviola, S., Carey, E., Mammarella, I. C., & Szucs, D. (2017). Stress, time pressure, strategy selection and math anxiety in mathematics: A review of the literature. *Frontiers in pPsychology, 8*, 1488. https://doi.org/10.3389/fpsyg.2017.01488

Chesney, D. L., McNeil, N. M., Brockmole, J. R., & Kelley, K. (2013). An eye for relations: eEye-tracking indicates long-term negative effects of operational thinking on understanding of math equivalence. *Memory & cCognition, 41*(7), 1079–1095. https://doi.org/10.3758/s13421-013-0315-8

Crooks, N. M., & McNeil, N. M. (2009). Increased practice with "set" problems hinders performance on the water jar task. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society, 31*(31), 643–648. https://philpapers.org/rec/CROIPW

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic mMedicine, 78*(8), 775–780. https://doi.org/10.1097/00001888-200308000-00003

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

De Caro, M. S. (2018). When does higher working memory capacity help or hinder insight problem solving? In Frédéric F. Vallée-Tourangeau (Ed.), *Insight: on the origins of new ideas* (pp. 79–104). Routledge.

DeCaro, M. S., Van Stockum, C. A., & Jr., & Wieth, M. B. (2016). When higher working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 39–49. https://doi.org/10.1037/xlm0000152

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review, 127*(3), 455–476. https://doi.org/10.1037/rev0000175

Duncker, K. (1945). On problem solving (L. S. Lees, Trans.). *Psychological Monographs, 58*(5 Whole No. 270), 1–113. https://doi.org/10.1037/h0093599

Ellis, J. J., & Reingold, E. M. (2014). The Einstellung effect in anagram problem solving: eEvidence from eye movements. *Frontiers in Psychology, 5*, 679. https://doi.org/10.3389/fpsyg.2014.00679

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two- factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). Elsevier.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Eemotion, 6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences, 50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Fischer, H., & Holt, D. V. (2017). When high working memory capacity is and is not beneficial for predicting nonlinear processes. *Memory & Ccognition, 45*(3), 404–412. https://doi.org/10.3758/s13421-016-0665-0

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology, 83*(3), 269–275. https://doi.org/10.1016/j.ijpsycho.2011.09.023

Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking & Reasoning, 15*(4), 355–376. https://doi.org/10.1080/13546780903178615

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika, 81*(3), 515–526. https://doi.org/10.1093/biomet/81.3.515

Haager, J. S., Kuhbandner, C., & Pekrun, R. (2014). Overcoming fixed mind sets: The role of affect. *Cognition & Emotion, 28*(4), 756–767. https://doi.org/10.1080/02699931.2013.851645

Huang, F., Zhao, Q., Zhou, Z., & Luo, J. (2019). People got lost in solving a set of similar problems. *NeuroImage, 186*(1), 192–199. https://doi.org/10.1016/j.neuroimage.2018.10.063

Imbo, I., & LeFevre, J. A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese-and

Canadian-educated adults. *Memory & Cognition, 38*(2), 176–185. https://doi.org/10.3758/MC.38.2.176

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science, 17*(3), 262–269. https://doi.org/10.1111/j.1467-9280.2006.01695.x

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*(4), 544–564. https://doi.org/10.1037/0022-3514.93.4.544

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies, 12*(1), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE, 7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: aA self-learning text* (3rd ed.). Springer. https://doi.org/10.1007/0-387-29150-4

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*, 1000–1009. https://doi.org/10.3758/BF03195762

Kobayashi, A., & Okubo, M. (2014). Assessment of working memory capacity with a Japanese version of the Operation Span Test. *Japanese Journal of Psychology, 85*(1), 60–68. https://doi.org/10.4992/jjpsy.85.60

Laird, B. K., Bailey, C. D., & Hester, K. (2018). The effects of monitoring environment on problem-solving performance. *The Journal of Social Psychology, 158*(2), 215–219. https://doi.org/10.1080/00224545.2017.1324396

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(6), 1–95. https://doi.org/10.1037/h0093502

Luchins, A. S., & Luchins, E. H. (1950). New experimental attempts at preventing mechanization in problem solving. *The Journal of General Psychology, 42*(2), 279–297. https://doi.org/10.1080/00221309.1950.9920160

Maule, A.J., Hockey, G.R.J. (1993). State, stress, and time pressure. In: Svenson, O., Maule, A.J. (eds) *Time Pressure and Stress in Human Judgment and Decision Making*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-6846-6_6

Minton, H. L. (1963). A replication of perceptual curiosity as a function of stimulus complexity. *Journal of Experimental Psychology, 66*(5), 522–524. https://doi.org/10.1037/h0044748

Neroni, M. A., Vasconcelos, L. A., & Crilly, N. (2017). Computer-based "mental Set" tasks: An alternative approach to studying design fixation. *Journal of Mechanical Design, 139*(7), 071102. https://doi.org/10.1115/1.4036562

Ninomiya, Y., Terai, H., & Miwa, K. (2022). Differences in the distribution of attention to trained procedure between finders and non-finders of the alternative better procedure. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.934029

Noël, M. P., Désert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory & Ccognition, 29*(1), 34–42. https://doi.org/10.3758/BF03195738

Ohlsson, S. (2018). The dialectic between routine and creative cognition. In Frédéric F. Vallée-Tourangeau (Ed.), *Insight: on the origins of new idea* (pp. 79–104). Routledge.

Okada, T., & Ishibashi, K. (2017). Imitation, inspiration, and creation: Cognitive process of creative drawing by copying others artworks. *Cognitive Sscience, 41*(7), 1804–1837. https://doi.org/10.1111/cogs.12442

Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental Set due to domain knowledge. *Memory & Cognition, 35*(6), 1456–1462. https://doi.org/10.3758/BF03193615

Ruiz, S., Chen, X., Rebuschat, P., & Meurers, D. (2019). Measuring individual differences in cognitive abilities in the lab and on the web. *PLOS ONE, 14*(12), e0226217. https://doi.org/10.1371/journal.pone.0226217

Sahai, R., & Frese, M. (2019). If you have a hammer, you only look for nails: tThe relationship between the Einstellung effect and business opportunity identification. *Journal of Small Business Management, 57*(3), 927–942. https://doi.org/10.1111/jsbm.12346

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs, 128*(2), 165–207. https://psycnet.apa.org/record/2002-15591-003. Accessed 2023/08/03

Sheridan, H., & Reingold, E. M. (2013). The mechanisms and boundary conditions of the Einstellung effect in chess: Evidence from eye movements. *PLOS ONE, 8*(10), e75796. https://doi.org/10.1371/journal.pone.0075796

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Sowden, P. T., Pringle, A., & Gabora, L. (2015). The shifting sands of creative thinking: Connections to dual-process theory. *Thinking & Reasoning, 21*(1), 40–60. https://doi.org/10.1080/13546783.2014.885464

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A, 58*(7), 1311–1323. https://doi.org/10.1080/02724980443000593

Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problems: The role of working memory capacities. *European Journal of Cognitive Psychology, 18*(5), 756–775. https://doi.org/10.1080/09541440500412270

Thevenot, C., & Oakhill, J. (2008). A generalization of the representational change theory from insight to non-insight problems: The case of arithmetic word problems. *Acta Psychologica, 129*(3), 315–324. https://doi.org/10.1016/j.actpsy.2008.08.008

Thomas, C., Didierjean, A., Maquestiaux, F., & Gygax, P. (2015). Does magic offer a cryptozoology ground for psychology? *Review of General Psychology, 19*(2), 117–128. https://doi.org/10.1037/gpr0000041

Thomas, C., Didierjean, A., & Kuhn, G. (2018). It is magic! How impossible solutions prevent the discovery of obvious ones? *Quarterly Journal of Experimental Psychology, 71*(12), 2481–2487. https://doi.org/10.1177/1747021817743439

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and lLanguage, 28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104.

Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 43(12), 1962–1987.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Rresearch Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720

Van Stockum, C. A., Jr., & De Caro, M. S., Jr. (2020). When working memory mechanisms compete: Predicting cognitive flexibility versus mental set. *Cognition, 201*, 104313. https://doi.org/10.1016/j.cognition.2020.104313

Visu-Petra, L., Miclea, M., Cheie, L., & Benga, O. (2009). Processing efficiency in preschoolers' memory span: Individual differences related to age and anxiety. *Journal of Experimental Child Psychology, 103*(1), 30–48. https://doi.org/10.1016/j.jecp.2008.09.002

Visu-Petra, L., Cheie, L., & Miu, A. C. (2013). Working memory and anxiety. In T. PackiamAlloway & R. G. Alloway (Eds.), *Working memory: The connected intelligence* (pp. 209–238). Psychology Press.

Wang, Z., & Shah, P. (2014). The effect of pressure on high-and low-working-memory students: An elaboration of the choking under pressure hypothesis. *British Journal of Educational Psychology, 84*(2), 226–238. https://doi.org/10.1111/bjep.12027

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science, 21*(4), 258–262. https://doi.org/10.1177/0963721412447622