Investigation of effects of working memory capacity on rule discovery process using eye movement data

Miki Matsumuro (muro@cog.human.nagoya-u.ac.jp)
Kazuhisa Miwa (miwa@is.nagoya-u.ac.jp)
Graduate School of Information Science, Nagoya University, Fro-cho, Chikusa-ku, Nagoya, Japan

Abstract

Many studies have investigated the process of rule discovery. However, the data utilized in these studies, such as performance and verbal protocol data, were course-grained. In this study, we designed a new experimental method using eye movement data to observe the detailed process of rule discovery. In the proposed method, we corresponded the task display and a rule space in the participants' minds to understand how they consider the rules and observe instances by eye tracking. Then, we compared the process of rule discovery by people with high and low working memory capacities. The results of the experiment revealed that those with high working memory capacity tried to consider one or similar rules from multiple instances. On the other hand, those with low working memory capacity tended to consider various rules from one instance.

Keywords: Rule discovery; eye tracking; working memory capacity; search strategy

Introduction

It is one of the most important activities to find regularities not only in science but also in many aspects of daily life. In this study, we tackle two goals in order to understand the process of rule discovery. Our first goal is to propose a new experimental method so as to observe the detailed process of rule discovery. The second goal is to investigate the relation between working memory capacity (WMC) and strategies of rule discovery.

The origin of this study is traced back to the dual space search theory proposed by Simon and Lea (1974). They suggested that the process of rule discovery develops through the interaction between two types of searches in two spaces, a rule space and an instance space. Problem solvers state rules by searching in a rule space while generating and observing instances in an instance space, and modify the rules or propose new rules based on their observations.

Although their theory can successfully explain the process of rule discovery, when conducting experiments based on this theory, there are limitations. One is that it is difficult to observe the detailed process of search in a rule space because the thought process to state rules cannot be monitored directly. To investigate this process, the researchers have utilized the protocol analysis method (e.g., Haverty, Koedinger, Klahr, & Alibali, 2000). The participants' think-aloud protocol data were used for this analysis. However, the data were coarse-grained on the time scale, and the participants mentioned only their conscious thoughts. Therefore, the analysis of the detailed patterns of search in the two spaces from protocol data often faced essential limitations. For these reasons, our first purpose is to propose a new experimental method to observe the detailed process of rule discovery.

The second goal is to investigate the relation between WMC and strategies of rule discovery. Dougherty and Hunter (2003) showed that people with high WMC maintained more alternative hypotheses in their mind when they engaged in hypothesis generation. However, their task was a probability judgment task and their analysis was based on participants' performance of the thought-listing task that was even more coarse-grained than the protocol data. Other studies also analyzed the effects of WMC based on the score of the reasoning tasks (e.g., Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). In the current study, we focus not on outputs as results but on the process of rule discovery by using eye movement data.

In summary, the first purpose of our study is to propose a new experimental method by which we obtain detailed data about the process of rule discovery. The second purpose is to investigate how participants' WMC affects such a process. Our method using eye tracking was established based on the SDDS (Scientific Discovery as Dual Search) model of Klahr and Dunbar (1988). The reason we use eye tracking is that eye movement data are more fine-grained than verbal reports and obtained directly without participants' conscious effort (cf. Rehder & Hoffman, 2005). With this method, we examine differences in search strategies affected by WMC, selecting those who have high or low WMC based on screening test scores.

Experimental Method

Search in Each Space

We designed an experimental method by means of eye tracking. We utilize the idea of the structure of a hypothesis space in the SDDS model. It extends the dual space search theory for investigating scientific discovery. The hypothesis space in the SDDS model corresponds to a rule space in dual space search. The hypothesis space includes all available hypotheses, each of which is connected with others through search paths. The search path does not connect hypotheses that have different schema. Therefore, similar hypotheses with an identical schema construct a subspace of the hypothesis space. Based on this idea, in our method, we manipulate the structure of a rule space by giving a different function to each rule. Since a different function evokes a different schema, only rules that have the same function are connected, and they construct a subspace in a rule space.

We obtain search patterns in the rule and instance spaces by means of eye tracking. First, we define search in each space as follows. The search in a rule space is the process where

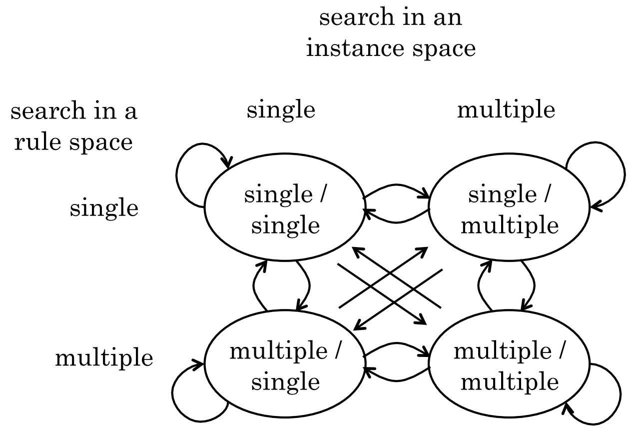

Figure 1: Four search statuses of the rule and instance spaces and transition patterns among the statuses.

participants consider rules from instances. On the other hand, the search in an instance space is defined as the activities where participants generate or observe instances to consider rules. In actual processes of rule discovery, timing to switch between searches in the rule and instance spaces is not clear. Therefore, in this method, we collect eye movement data as the search in a rule space and observations of instances as the search in an instance space.

There are two modes in the search process (Figure 1): the *single-* and *multiple-subspace search* in a rule space and the *single-* and *multiple-instance search* in an instance space. First, we define two modes of search in the rule space. The single-subspace search is the search mode in which participants search only in one subspace. In other words, they consider rules in the same subspace that are characterized by an identical function. On the other hand, in the multiple-subspace search mode, they try to decide which subspace they should search. Whereas the single-subspace search is a "search in a subspace," the multiple-subspace search is a "search for a subspace" to be focused on.

For search in the instance space, we also define two modes of search. Participants in the single-instance search mode focus on observing one instance. They consider multiple rules from a single instance. In contrast, they observe various instances for stating rules in the multiple-instance search mode. They compare multiple instances to obtain cues for rule discovery.

Each participant's search status is categorized into one of the four search statuses in Figure 1. The single/single status is one in which participants observe one instance and consider rules with an identical function. When participants consider a complex process, they do so in this status because they may concentrate their attention on a single event. Participants in the single/multiple status also consider rules in an identical subspace, but they observe and compare multiple instances to obtain the cues for rule discovery. On the other hand, in the multiple/single status, participants consider rules with various functions across multiple subspaces while focusing on examining one instance. Last, the participants in the multiple/multiple status also consider various rules across multiple subspaces while observing multiple instances in a short time. This status seems not to be suitable for rule searching because they simultaneously vary both factors, rules and instances, preventing efficient searches for rule discovery. To identify each search status (a circle in Figure 1) and each transition among the search statuses (an arrow in Figure 1), we systematically designed a rule discovery task as follows.

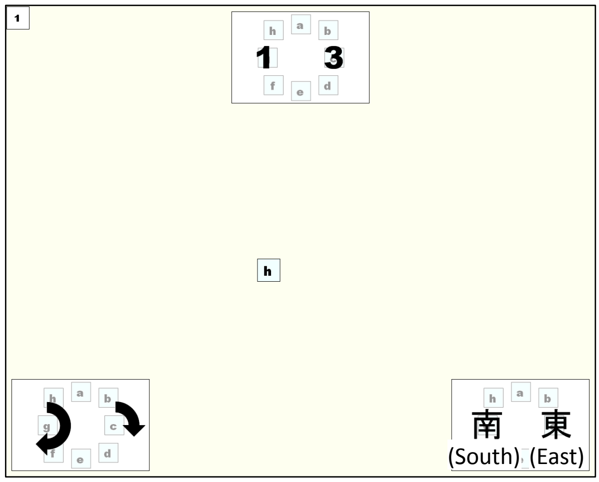

Figure 2: Example screenshot of task display in this study. The upper panel is the number panel, the lower-left panel is the arrow panel, and the lower-right panel is the compass panel. The directions were presented in Chinese characters. The target alphabet is presented in the center of the display.

Rule Discovery Task Using Eye Tracking

The task display consists of three panels (arrow, compass, and number panels) and eight letters (a to h) in the center of the display as shown in Figure 2. The eight letters are placed in circle. The presented letter (we call this the "target") is determined based on the objects displayed on one of the three panels. The participants are asked to find a rule determining which panel relates to the target and how the target is determined by the objects in the related panel. Only one of the three panels relates to the target and the other two panels have no relation to it. The participants are required to find a rule as quickly as possible.

The experimental procedures are as follows. Before the experiment, an experimenter prepares six instances as stimuli. In the example instance in Figure 2, the rule may be: "the target is the letter (h) in the opposite position (North-West) of the combination of two directions (South-East) on the compass panel." In this case, the experimenter selects five other instances so as to be instances consistent with the same compass rule. The participants are instructed that only one panel is related to the target in advance. The participants observe all six instances one-by-one using right- and left-arrow keys on the keyboard. The panel that has the same function is always presented in the same position on the display. Additionally, the six instances are presented in a cyclic manner; the participants can observe instances as many times as they want. When the participants think of a rule, they press the space key and report their rule to the experimenter. When the

Table 1: Functions and example rules of each panel

panel	function	how to decide target	example rule (see example in Figure 2)
number	order	a letter corresponds to a number in alphabetical order ("a" is 1)	Target is the sum of numbers in alphabetical order (e.g., 1 + 3 = 4 → d)
arrow	rotation	a letter corresponds to an angle of arrows	Target is a shifted letter from "a" by an angle of left object (e.g., 180 degrees → e)
compass	position	a letter corresponds to a direction (based on map)	Target is pointed at by right object (e.g., East → c)

correct rule is discovered, the experiment is terminated. Every five minutes, the calibration for recording eye movement is performed. Table 1 shows functions and example rules in each panel. Before starting the task, the participants learn these functions sufficiently.

Data and Search in Each Space

Figure 3 shows example data obtained in the experiment. For example, from 520 to 530 seconds in Figure 3, this participant focused her attention on the arrow panel, meaning that she considered only rules that use arrows; that is, she searched in the subspace of "rotation" rules in a rule space. The shift of fixation from the arrow to the number panel at around 530 seconds means that her search subspace shifted to the "order" rule subspace. In the same manner, the behavior after 530 seconds is interpreted that she expanded her attention toward the three panels, meaning that she searched all subspaces in a rule space simultaneously. Soon after broadening the search, she fixed her attention on the compass panel. In the final part, she shifted her search subspace to the "order" rule without broadening the search. On the other hand, for the search in an instance space, the participant seemed to focus on a single instance before around 560 seconds. Then, she moved to observe multiple instances during a short time after around 570 seconds.

With reference to Figure 3, around 520 seconds, her search mode was in the single/single status. Then, she searched in multiple subspaces in a rule space with an identical instance at around 530 seconds, meaning that her search status shifted to the multiple/single. Soon, she came back to the single/single status, and this status continued for a while. After around 570 seconds, she observed multiple instances one-by-

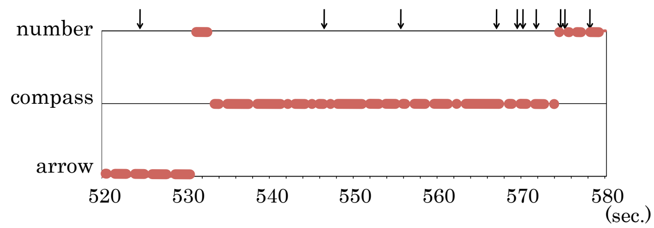

Figure 3: Timeline graph of collected data.
Each red horizontal line indicates a panel that the participant focused on. Labels of the vertical axis indicate three panels. Arrows on the top of the graph mean timing at which the participant shifted an instance to the previous or next one.

one, meaning that she moved to the single/multiple status. As shown in this example, it is possible to obtain detailed search processes in the rule and instance spaces by our experimental method.

We conducted an experiment with this method to observe the process of searches in the rule and instance spaces to understand the effect of WMC on the rule discovery process.

Experiment

First, we measured the participants' spatial span capability as a screening test.

Screening Test

Fifty-seven undergraduates engaged in the spatial span test by Shah and Miyake (1996). In the task, a capital letter was presented on a computer screen that was not in the upright orientation. The participants were asked to indicate whether the letter was normal or mirror-imaged and were required to store the orientation of each letter for a subsequent recall test. One participant whose judgment score was around the chance level was excluded from analysis. Fifty-six participants' data were scored based on the partial-credit unit scoring by Conway et al. (2005). The mean score was 0.576 ($SD = 0.200$), the highest score was 0.920, and the lowest score was 0.121.

Rule Discovery Task

Method

Participants The twelve participants with the highest WMC scores participated as the high-WMC group. Their mean score was 0.825 ($SD = 0.068$). The twelve participants with the lowest WMC scores participated as the low-WMC group. Their mean score was 0.290 ($SD = 0.091$). The mean WMC score of the high group was significantly higher than that of the low group ($t(22) = 16.318, p < .001$).

Apparatus We presented the task display on a 17-in. monitor with a resolution of 1280 × 1024 pixels. The participants were seated approximately 60 cm away from the monitor. The size of panels was approximately 5.25° × 7.82° of visual angle. Each panel was placed as shown in Figure 2 on the upper-center, lower-left, and lower-right of the task display. The participants' eye movements were recorded using the Tobii T60 eye tracker at 60 Hz. The participants were allowed to move their heads naturally.

Table 2: Rules used in this study

	panel	rule
task 1	compass	Target is the opposite position of combination of two directions on the panel
task 2	arrow	Target is a shifted letter by sum of an angle of two arrows on the panel and 135 degrees
task 3	number	Target is the sum of the difference between two numbers and the bigger one on the panel

Procedures Approximately one month after the screening test, the rule discovery task was conducted individually. First, the participants were instructed on the task and learned the function of each panel sufficiently through practice. Task 1 was preliminarily performed to let the participants establish their own strategies. In task 1, the participants needed to find a simple rule in ten minutes. Soon after task 1 ended, task 2 was conducted with the same procedure as that in task 1. The participants needed to find a relatively complex rule in twenty minutes in task 2. The data of task 2 were used for analysis. As an exception, the participants who found the rule before fifteen minutes passed were led to engage in another task (task 3) with the same procedure as that in task 2. In this case, the data of task 3 were used for analysis. Table 2 shows the rules used in the experiment.

Results

One participant in the low group was excluded from analysis because we did not obtain the data for more than fifteen minutes. We analyzed only the participants from whom more than 60% of eye movement data were recorded correctly. Based on the criterion, six participants in the high group and two participants in the low group were excluded; therefore, the data of six participants in the high group (WMC score $M = 0.858, SD = 0.058$) and nine participants in the low group (WMC score $M = 0.289, SD = 0.100$) were used for analyses. The average WMC score of the high group was still significantly higher than that of the low group ($t(13) = 12.551, p < .001$).

Search Status Fixations longer than 100 msec were analyzed, and fixations outside the panels were excluded from analysis. First, we traced each participant's data along a timeline to acquire on which panel the fixations were observed and when an instance was shifted, as shown in Figure 3. We defined that the fixation shift happened when the participant's fixation point was observed on a different panel from the preceding one. Similarly, we defined that the instance shift happened when the participants pressed the arrow keys to observe another instance. Based on the median of the fixation time in one panel (8.961 seconds) and the mean of observation time in one instance (7.531 seconds), we segmented the timeline every 7.5 seconds.

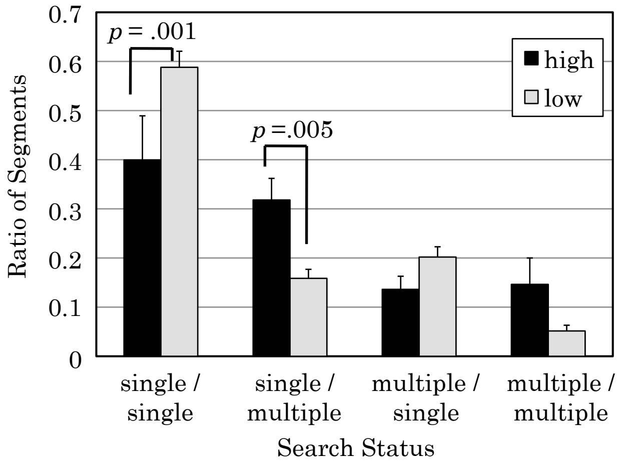

Figure 4: Ratio of segments in four search statuses in each WMC group (bars show standard errors).

Then, we counted how many fixation and instance shifts happened in each segment. We also performed the same analysis with the segmentations at six, seven, and eight seconds, and similar results were confirmed. We defined that the participants were in the multiple-subspace search mode in a rule space when the fixation shifts were observed more than once in each segment. Similarly, for search in an instance space, in each segment when more than one instance shift were observed, the participants were defined to be in the multiple-instance search mode. Based on the definitions, we classified the participants' status in each segment into four categories shown in Figure 1.

Figure 4 presents the ratio of segments categorized into each search status in each WMC group. Mixed ANOVA with WMC (high and low) as between-subject factor and search status (four statuses) as within-subject factor was performed on the ratio of segments. The interaction between the WMC and the search status reached significance ($F(3,39) = 6.328, p = .001$). The ratio in the single/single status in the high-WMC group was significantly higher than that in the low-WMC group ($F(1,52) = 12.116, p = .001$). On the other hand, the ratio in the single/multiple status in the high-WMC group was significantly lower than that in the low-WMC group ($F(1,52) = 8.663, p = .005$). These results show that the participants in the high-WMC group compared multiple instances more frequently than those in the low-WMC group when they were in the single-subspace search mode in a rule space search. On the other hand, the participants in the low-WMC group searched in the single subspace in a rule space with fewer shifts of instances. When they searched in a rule space with the multiple-subspace search mode, there was no significant difference in search status between the two WMC groups.

Transition Probability among Each Search Status Next, we calculated the transition probability of shifting from one to another or being in the same search status. There are sixteen transition patterns shown in Figure 1. Figure 5 shows

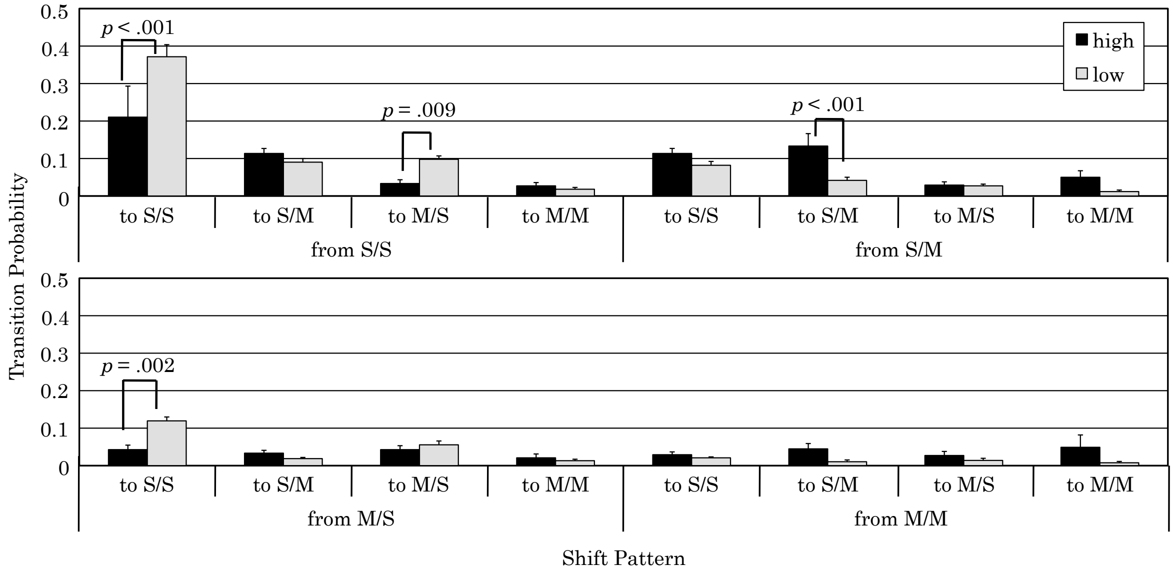

Figure 5: Probability of each transition pattern in each WMC group (bars show standard errors).
S/S = single/single, S/M = single/multiple, M/S = multiple/single, M/M = multiple/multiple.

the transition probability of each transition pattern in each WMC group. Mixed ANOVA with WMC (high and low) as between-subject factor and transition pattern (sixteen patterns) as within-subject factor was performed on the transition probability. The interaction between the WMC and the transition pattern reached significance ($F(15,195) = 5.323, p < .001$). Four comparisons reaching significance are shown in Figure 5: the transition probability of single/single to single/single in the low-WMC group was significantly higher than that in the high-WMC group ($F(1,208) = 43.283, p < .001$); the transition probability of single/single to multiple/single in the low-WMC group was significantly higher than that in the high-WMC group ($F(1,208) = 6.955, p = .009$); the transition probability of single/multiple to single/multiple in the high-WMC group was significantly higher than that in the low-WMC group ($F(1,208) = 14.012, p < .001$); and the transition probability of multiple/single to single/single in the low-WMC group was significantly higher than that in the high-WMC group ($F(1,208) = 9.819, p = .002$). Other comparisons did not reach significance.

These results indicate that when they searched in a single subspace of a rule space, the participants in the high-WMC group tended to continue the multiple-instance search in an instance space, and those in the low-WMC group tended to continue to the single-instance search. On the other hand, the participants in the low-WMC group tended to shift to the search in multiple subspaces of a rule space focusing on a single instance more frequently than those in the high-WMC group, even though their multiple-subspace search did not continue very long.

Discussion

Our first purpose was to propose a new experimental method by which we could capture the detailed process of rule discovery. This purpose was achieved by manipulating a structure of a rule space and using eye tracking. We successfully observed the search statuses in the rule and instance spaces and the transition patterns among those in detail. The second purpose of this study was to compare the process of rule discovery by people with high or low WMC. Our method detected the different processes of the two types of participants.

Advantages of New Method

The protocol analysis has two main limitations; as a result, we could not capture the detailed process of search in a rule space. The first difficulty is the grain size of data. Reporting one's thoughts cannot happen at the speed of thought, making it impossible for participants to report all of their thoughts. Participants would omit reporting their short-term thoughts. The use of eye tracking has the potential to solve this problem. We can directly obtain which direction participants focus their attention in by eye tracking. Furthermore, the sampling rate is very fine, 60 Hz in this study.

The second limitation of verbal protocols appears when participants have difficulty putting their thoughts into words. Participants sometimes have "no idea" during an experiment. In this study, all except one participant reported in the post-task interview that there were periods when they came up with no idea. In such a situation, the participants were usually confused and had difficulties in putting their thoughts into words. Moreover, the verbalization of confused thoughts would further muddle their thoughts. Eye movement data are always recorded throughout the task, even if participants cannot verbalize their thoughts.

We designed our task based on the SDDS model by Klahr and Dunbar (1988) to maximize the advantages of eye movement analysis. The task display as the observable externalized space consistently corresponds to the subspaces in a rule space as the internal representation. Due to this design, we could analyze search in a rule space from eye movement data. These features in our experimental method enabled us to capture each search status and the transition among the statuses in detail.

Differences between Each WMC Group

As a result of the experiment, different strategies were observed according to participants' WMC.

The analysis of the occurrence ratio of each search status and each transition pattern suggested that the participants in the high-WMC group tended to search in a rule space focusing on one subspace compared to those in the low-WMC group. Additionally, when they considered the rules that have an identical function, they observed and compared multiple instances more continuously. These results suggest that the participants with high WMC considered complex rules with an identical function while comparing multiple instances, meaning that they preferred the depth-first search in a rule space. Note that they did not necessarily fix their search to one single subspace because they actually shifted the search from one to another subspace at their own pace.

On the other hand, quick shifts of fixation among multiple panels, i.e. being in the multiple-subspace search mode in a rule space, were more frequently observed in the low-WMC group. However, this mode did not continue, and they soon came back to the single-subspace search mode. These results indicate that the participants with low WMC searched in a rule space switching the single- and multiple-subspace search modes alternatively. For the search in an instance space, they did not observe multiple instances as the participants in the high-WMC group did. This implies that they tried to consider rules from one instance, whereas the participants in the high-WMC group tried to consider one rule from multiple instances. These results suggest that they valued the breadth-first search. These two strategies, depth-first and breadth-first searches, were also observed in our previous study (Matsumuro & Miwa, 2011).

We suggest a possible reason that the search strategy was different according to their WMC based on the studies of category learning. Two strategies have been shown (DeCaro, Thomas, & Beilock, 2008; Rehder & Hoffman, 2005): the rule-based strategy where participants conduct hypothesis testing explicitly, and the information-integration strategy where rules are learned implicitly by integrating stimuli across multiple dimensions. The participants in the high-WMC group in our study observed multiple instances while searching in a single subspace in a rule space. They would generate hypotheses by comparing instances and tested these hypotheses as with the rule-based strategy. The explicit strategy was suitable for the participants with high WMC because it relies heavily on working memory. By contrast, the participants in the low-WMC group searched in multiple subspaces with a single instance. They would have gathered information from multiple dimensions and tried to figure out the relation between the target and the objects on the three panels implicitly, as with the information-integration strategy. The implicit strategy is processed without conscious control. Therefore, this strategy is suitable for the participants with low WMC because it does not require working memory load. Given these points, it may be possible that our participants in each group selected a strategy suitable for each WMC. Note that the search in multiple subspaces by the participants in the low-WMC group did not continue. They had to conduct explicit hypothesis testing because all participants were required to find a rule that could be reported in words.

To investigate this possibility, we should analyze verbal protocols along with the eye movement data recorded in this study. Additionally, in future work, we need to investigate the relation between the search strategies and the discovery rate, and interaction of such a relation and individual differences.

References

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12(5)*, 769–786.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107(1)*, 284–294.

Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, *113(3)*, 263–282.

Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving inductive reasoning problems in mathematics: Not-so-trivial pursuit. *Cognitive Science*, *24(2)*, 249–298.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12(1)*, 1–48.

Matsumuro, M., & Miwa, K. (2011). An investigation of search strategies for hypothesis generation using eye movement data. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 3424–3429). Boston, MA: Cognitive Science Society.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51(1)*, 1–41.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125(1)*, 4–27.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, NJ: Lawrence Erlbaum.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability - and a little bit more. *Intelligence*, *30(2)*, 261–288.