Human Discovery Processes Based on Searching Experiments in Virtual Psychological Research Environment

Kazuhisa Miwa

Graduate School of Human Informatics, Nagoya University, Nagoya 464-8601, Japan
miwa@cog.human.nagoya-u.ac.jp

Abstract. For designing experiments in social and human sciences, we must often consider various complex factors that seem to decide subjects' performance. It is sometimes difficult to make complete experimental planning in which hypotheses guiding the experiments are established prior to executing the experiments. Even if the situation stands, experts in the field systematically organize their experimental processes. We propose Searching Experimental Scheme (SES) that enables them do so. For confirming the validity of SES, we construct virtual psychological experimental environment using a cognitive simulator in which subjects try to generate hypotheses and conduct experiments as scientists do. We analyze the subjects' behavior based on SES and discuss the relation between the characteristics of their behavior and their performance of discovering targets.

1 Introduction

We can divide the ways of acquiring empirical data in the process of discovery into two basic categories: experimentation and observation. In experimentation, data are systematically gathered based on previously formed hypotheses. In experimental psychology, the most orthodox example is Factorial Design (FD) experiments in which focused factors by a researcher are systematically manipulated by clearly established hypotheses, and the relation between the manipulated factors and the observed data is identified. On the other hand, in observation, the systematic data collection as above is not made. In usual cases, hypotheses for manipulating experimental factors cannot be formed. So ways of gathering data become Trial and Error (TE) search in which experimental data are randomly observed for forming an initial hypothesis.

Experimental design in real research environment usually reflects the characteristics of both of the two typical categories above. For example, searching some levels of certain controlled factors may be lost even though the global structure of the experimental design is FD; or the experimental design is locally FD but the global structure (i.e. the relation of each local unit) may be TE. We call these intermediate ways of experimentation "searching experiments", which is a key concept of this study. The process of *searching experiments* appears when (1) an experimental space that subjects try to search is huge, so the subjects cannot search whole combinations of the experimental conditions at once, (2) a goal itself is ambiguous, that is, a research objective itself is being searched, and (3) the relation between independent and dependent factors cannot be clearly predicted because of the lack of knowledge on the research domain or the existence of complex interaction among the experimental factors.

Searching experiments are essentially important especially in social and human sciences because most of the research situations are relatively complex and satisfy the conditions above [1] [7]. Researchers use *searching experiments* effectively for organizing their experimental processes systematically under the complex research situations. In this study, we propose "Searching Experimental Scheme" that enables researchers perform systematic search even though well-organized experimental planning such as FD experiments cannot be adopted. Then we analyze subjects' behavior based on the scheme. We also discuss the relation between the characteristics of searching behavior and the performance of subjects' discovering targets. To do so, using a discovery task that satisfies the conditions in which *searching experiments* appear, we let subjects experience a series of experimental processes, such as setting up a research objective, forming a hypothesis, designing experiments, performing experiments, interpreting experimental results, and rearranging additional experiments.

To discuss the issues above, it is difficult to let subjects perform real psychological experiments because of its executing cost. So in this study we let them perform virtual psychological experiments using a cognitive simulator that is constructed as a computer program instead of performing real experiments. Subjects behave as an experimental psychologist in the research environment provided by the simulator [8].

2 Virtual Psychological Research Environment

2.1 Wason's 2-4-6 Task

The simulator used in this study is a cognitive model that simulates collaborative discovery processes in which two problem solvers interactively solve a traditional discovery task, the Wason's 2-4-6 task, while referring mutual experimental results [9]. Subjects participate in this experiment as an experimental psychologist who studies collaborative discovery processes using the Wason's task.

The standard procedure of the 2-4-6 task is as follows. Subjects are required to find a rule of relationship among three numerals. In the most popular situation, a set of three numerals, "2, 4, 6", is presented to subjects at the initial stage. The subjects form hypotheses about the regularity of the numerals based on the presented set. Subjects conduct experiments by producing a new set of three numerals and present them to an experimenter. This set is called an instance. An experimenter gives Yes feedback to subjects if the set produced by subjects is an instance of the target rule, or No feedback if it is not an instance of the target. Subjects carry out continuous experiments, receive feedback from each experiment, and search to find the target.

Two types of experimentation, Ptest and Ntest, are considered. Ptest is experimentation using a positive instance for a hypothesis, whereas Ntest is experimentation using a negative instance. For example, when a subject has a hypothesis that three numerals are evens, an experiment using an instance, "2, 8, 18", corresponds to Ptest, and an experiment with "1, 2, 3" corresponds to Ntest. Note that the positive or negative test is defined based on a subject's hypothesis, on the other hand, Yes or No feedback is on a target. We should also notice the pattern of hypothesis reconstruction based on the combination of a hypothesis testing strategy and an experimental result (Yes or No feedback from an experimenter). When Ptest is conducted and No feedback is given, the hypothesis is disconfirmed. Another case of disconfirmation is the combination of Ntest and Yes feedback. On the other hand, the combinations of Ptest - Yes feedback and Ntest - No feedback confirm the hypothesis.

2.2 Interactive Production System

We have developed an interactive production system architecture for constructing the cognitive simulator and providing the virtual psychological research environment. The architecture consists of five parts: production sets of System A; production sets of System B; a working memory of System A; a working memory of System B; and a commonly shared blackboard (see Figure 1). The two systems interact through the common blackboard. That is, each system writes elements of its working memory on the blackboard and the other system can read them from the blackboard. The model solving the Wason's 2-4-6 task has been constructed using this architecture.

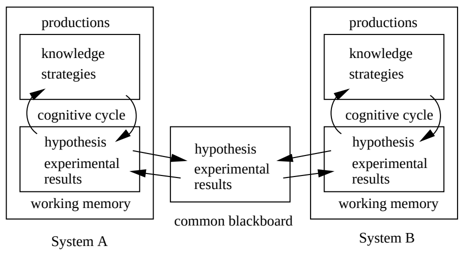

Fig. 1. Basic structure of the interactive production system architecture

The model has the knowledge on the regularities of three numerals, which is used for hypothesis generation in the process of solving the 2-4-6 task. The knowledge is organized as the dimension-value lists. For example, "continuous evens", "three evens", and "the first numeral is even" are example values of a dimension, "Even-Odd". The dimensions the systems use are: Even-Odd, Order, Interval, Range of digits, Certain digit, Mathematical relationship, Multiples, Divisors, Sum, Product, Different.

The way of searching the hypothesis space is controlled by the system's parameter that decides the hypothesis formation strategy (see Table 2).

Basically the model searches the hypothesis space randomly in order to generate a hypothesis (when a value of the parameter, [random], is set). Three hypotheses, "three continuous evens", "interval is 2", and "three evens", are particular. Human subjects tend to generate these hypotheses at first when the initial instance, "2, 4, 6", is presented. So our model also generates these hypotheses first prior to other possible hypotheses when a value of the parameter, [human], is set.

You can see the detailed specifications of this model in Miwa & Okada, 1996 [5].

2.3 An Example Behavior of the Simulator

Table 1 shows an example result of the computer simulations. The target was "Divisor of three numerals is 12". Two systems interactively found the target. One system, System A, always used Ptest in its experiments, and the other, System B, used Ntest. The table principally consists of three columns. The left-most and right-most columns indicate hypotheses formed by System A and System B respectively. The middle column indicates experiments, that is, generated instances, Yes or No feedback, and the distinction of Ptest or Ntest conducted by each system. Each experiment was conducted alternately by two systems, and the results of the experiments were sent to both of the two systems. The left-most number in each column indicates a series of processing, from #1 through #41. The right-most number in the left-most and right-most columns indicates the number of each hypothesis being continuously confirmed. System A disconfirmed its hypotheses at #4, #10, #16, which were introduced by self-conducted experiments at #3, #9, #15. System B disconfirmed its hypotheses at #17, #29, which were introduced by other-conducted experiments at #15, #27.

What we should note is that the simulator actually simulates human discovery processes. The validity of the simulator as a cognitive model has been already verified in other our papers. So usage of this simulator provides more realistic research environment in which we can observe searching processes of subjects who behave as an experimental psychologist.

2.4 Parameters Deciding the Model's Behavior

The parameters that decide the simulator's behavior consist of 6 factors that are indicated in Table 2. Five parameters except a first parameter, Target, are set up for controlling each of two interacting systems.

Let us now compare the situation in which *searching experiments* appear (see 1.) and the virtual experimental environment provided by the simulator. First, it is impossible to search the whole experimental space because it consists of two hundred million conditions ($= 35 \times 5^2 \times 4^2 \times 5^2 \times 5^2 \times 5^2$). Second, focused factors, which are decided based on research objectives, are independently selected by subjects themselves; actually achieved solutions of every subject are different.

Table 1. An example behavior of the simulator

Hypotheses by System A			Experiments				Hypotheses by System B		
			1	2, 4, 6	Yes				
2	Continuous evens numbers.	0	3	4, 6, 8	No	Ptest by SysA		-	
4	The product is 48.	0	6	6, 6, -17	No	Ntest by SysB	5	The sum is a maltiple of 4.	0
8	The product is 48.	1	9	24, -1, -2	No	Ptest by SysA	7	The sum is a maltiple of 4.	1
10	First + Second = Third.	0	12	3, -8, -20	No	Ntest by SysB	11	The sum is a maltiple of 4.	2
14	First + Second = Third.	1	15	-10, 2, -8	No	Ptest by SysA	13	The sum is a maltiple of 4.	3
16	Divisor is 12.	0	18	-5, -14, -9	No	Ntest by SysB	17	The second is 4.	0
20	Divisor is 12.	1	21	2, 4, 6	Yes	Ptest by SysA	19	The second is 4.	1
22	Divisor is 12.	2	24	-17, 3, 12	No	Ntest by SysB	23	The second is 4.	2
26	Divisor is 12.	3	27	2, 12, -12	Yes	Ptest by SysA	25	The second is 4.	3
28	Divisor is 12.	4	30	8, 12, -2	No	Ntest by SysB	29	Divisor is 12.	0
32	Divisor is 12.	5	33	2, 6, -2	Yes	Ptest by SysA	31	Divisor is 12.	1
34	Divisor is 12.	6	36	-2, -7, -8	No	Ntest by SysB	35	Divisor is 12.	2
38	Divisor is 12.	7	39	4, 3, -12	Yes	Ptest by SysA	37	Divisor is 12.	3
40	Divisor is 12.	8					41	Divisor is 12.	4

Third, there are complex interactions especially among three factors: hypothesis testing strategies, hypothesis formation strategies, and targets. These points support that the research environment used in this study embodies the situation in which *searching experiments* appear.

3 Experiments

Six graduate students participated in the experiment. They attended a graduate school class given by the author. The topic of the class was experimental psychological studies on human hypothesis testing. So the experimental situation was that the subjects who had obtained basic psychological knowledge on human hypothesis testing were required to study collaborative discovery processes in the experimental environment, applying the basic knowledge to the collaborative situation. Each subject individually participated in the experiment. After instructional guidance for 20 minutes, the main experiment, in which the subjects manipulated the simulator and studied collaborative discovery processes, was carried out for 2 hours; then an interview for 15 minutes was conducted.

In the main experiment, the subjects performed experiments manipulating the simulator independently. An experimental planning sheet was used; the sheet consisted of 5 items: (1) a research objective (what do they investigate), (2) hypotheses, (3) an experimental design, (4) experimental results, and (5) interpretation of the experimental results. The subjects filled out the former three items, and then they actually conducted experiments manipulating the simulator. After the experiments they filled out the latter two items of the sheet. They repeated this procedure during the main experiment. In the interview after the experimental session, subjects' conclusions (i.e. what do they find) through the whole experiments were identified.

Table 2. Six factors of the simulator

Factors	Levels
Target [T]	[1] - [35] Thirty-five kinds of targets that were used in the experiment. For example, Target #1 is "ascending numbers"; Target #35 is "three different numbers".
Hypothesis testing strategies [HT]	[0], [25], [50], [75], [100] The probability of conducting positive tests in generating instances. [100] and [0] mean that systems always conduct positive tests and negative tests, respectively.
Hypothesis formation strategies [HF]	[human], [random], [specific], [general] [human] means that systems generate hypotheses as humans do. [random]: generating hypotheses randomly. [specific]: generating specific hypotheses prior to general ones. [general]: generating general hypotheses prior to specific ones.
# of activated instances [AI]	[all], [6], [5], [4], [3] The number of instances that can be activated at once in the working memory when generating hypotheses.
# of maintained hypotheses [RH]	[all], [5], [4], [3], [2] The number of rejected hypotheses that can be maintained in the working memory.
Condition for terminating the search [TE]	[all], [5], [4], [3], [2] The number of continuos confirmations when systems terminate the search. [2] means when a hypothesis is continuously confirmed two times, systems recognize the hypothesis as the solution, and terminate the search.

After the combination of every level of the 6 factors is decided, twenty simulations are automatically executed in the condition. Then the experimental system presents (1) the ratio of each of the two systems correctly finding a target, (2) the ratio of at least one of the two systems reaching a correct solution, and (3) the average number of generated instances for reaching a correct solution. The system also presents a model's solution process of each simulation in addition to the final results as above; however, on the basis of the experimenter's instruction, the subjects only focus on the final performance of the systems and try to find factors that explain the performance. The experimental system automatically records subjects' experimental behavior. Additionally the processes were also recorded by a video camera, and subjects' verbal protocols were gathered. Those protocols were used as secondary data for identifying subjects' behavior when their description on the experimental planning sheet was ambiguous.

4 Searching Experimental Scheme

4.1 Expanded Search Within/Out of Focused Factors

In this study, we describe subjects' experimental processes based on "Searching Experimental Scheme" (SES). Figure 2 shows the experimental space consisting of the 6 factors of the simulator, that is, the combinations of every level of each factor [2]. From Factor 2 through Factor 6, every level is decided in each of the two systems. The bold lines of Figure 2 show an example combination: Factor 1, a used target is “ascending numbers” (Target #1); Factor 2 and Factor 3, the combination of hypothesis testing and formation strategies is positive testing and specific formation strategies in one system v.s. negative testing and general formation strategies in the other system; Factor 4, whole instances in the working memory can be activated; Factor 5, every hypothesis can be maintained in the memory; Factor 6, search is terminated when a hypothesis is supported by three continuous confirmations.

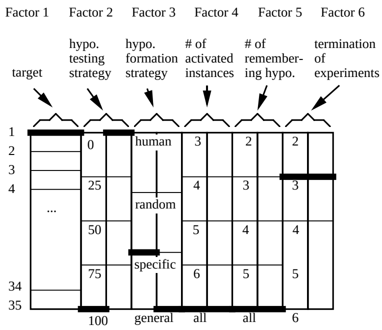

Fig. 2. experimental space of the simulator

Subjects’ behavior will be described on SES shown in Figure 3. Figure 3 consists of three basic units, Unit A11, Unit A12, and Unit A21. Each unit corresponds to a set of subjects’ searching behavior. We regard a series of continuous experiments guided by single experimental design on a piece of experimental planing sheet as a set of searching behavior.

In Unit A11, a subject manipulates Factor n and Factor m, and searches some levels of the factors indicated by the bold lines. We call these manipulated factors “focused factors”. Focused factors are indicated by dark gray boxes. Next in Unit A12, another factor, Factor p, indicated by a light gray box, is manipulated while fixing the levels of the focused factors already searched in Unit A11. We call this searching behavior “expanded search out of focused factors”.

Moreover, subjects do not necessarily search whole levels of focused factors within a single unit; so they often conduct additional search of the focused factors. For example, in Unit A21, a subject searches other levels of the focused factors than the levels that have been already searched in Unit A11. We call this searching behavior “expanded search within focused factors”. Although subjects cannot search all levels of focused factors at once because of their cognitive resource constraints, they try to analyze the effects of the focused factors on the total performance by conducting the expanded search within focused factors. Moreover, by conducting the expanded search out of focused factors, they try to

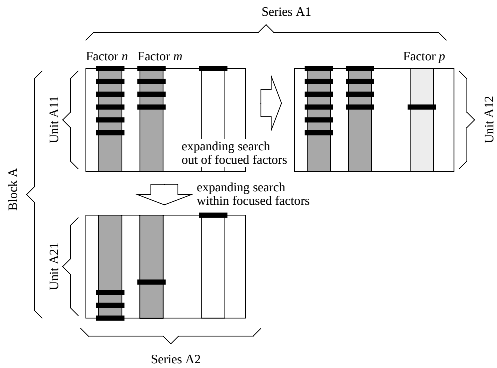

Fig. 3. Searching experimental scheme

know how the results obtained on the focused factors are affected by fluctuation of other factors. The expanded search within/out of focused factors reflects the characteristics of well organized *searching experiments*.

4.2 Levels of Searching Behavior

An important point in using SES defined above is that we can identify several levels of regularity of subjects' searching behavior. First, on the most basic level, a chunk (a unified set of subjects' searching behavior) is represented as each of Unit A11, Unit A12, and Unit A22. We call each chunk a "Unit". Next on the second level, a chunk is constructed by expanded search out of focused factors. We call this chunk a "Series". On this level, the subjects' behavior in Figure 3 is unified into two chunks: one chunk is Series A1 that consists of Unit A11 and Unit A12 and the other is Series A2. Finally on the third level, whole subjects' behavior in Figure 3, organized by expanded search within focused factors, is regarded as one chunk. We call this chunk on the highest level a "Block".

Now we should define each termination of a Series and a Block. A Series continues when subjects manipulate other factors than focused factors while fixing the already searched levels (or a part of the levels) of the focused factors. A Series terminates when conducting expanded search out of focused factors while shifting the search of the focused factors to new levels that have not been examined. A Block continues when subjects manipulate focused factors while fixing the already searched levels (or a part of the levels) of other factors than the focused factors. A Block terminates when both of focused factors and other factors are manipulated at once.

Figure 4 shows an example searching behavior of Subject B described based on SES.

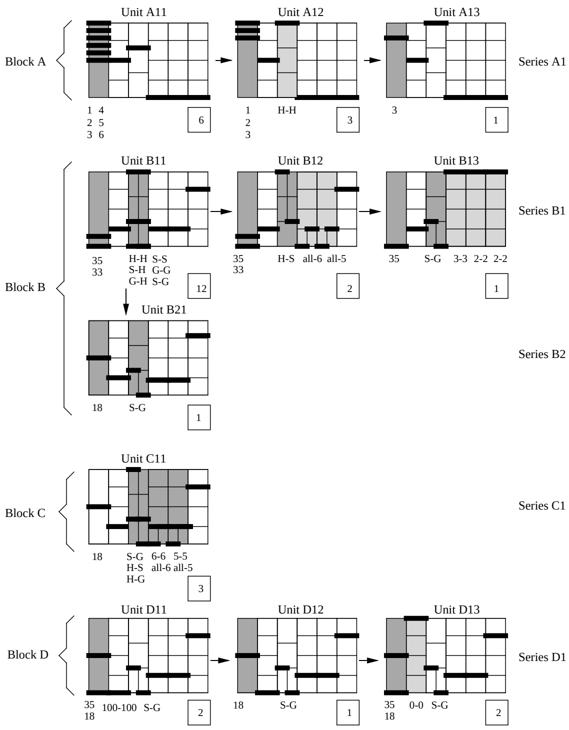

Fig. 4. An example behavior of Subject B described based on SES

4.3 Three Stages of Chunking

Figure 5 shows the total number of experiments of each subject and the numbers of chunks on the three levels (Unit, Series, and Block) defined in the previous section. Constructing chunks on the higher levels means higher organization of subjects' searching behavior; so Figure 5 indicates the situation of phased organization processes of subjects' searching behavior.

Now to model the patterns of the phased organization process, let us consider the 2 factorial (3×3) experimental design. Figure 6 (a) shows the case in which the experiments are performed based on FD in which all levels of two focused factors are searched at once. In this case a Unit is equal to a Block. Searching behavior is only organized through the process of constructing a Unit from individual experiments. We call this organization process "first stage of chunking". On the other hand, when experiments are performed based on TE search, every experiment is independent from each of the former and latter experiments; so no chunking happens. In this case each single experiment constructs a Block (see Figure 6 (e)). Characteristics of subjects' behavior of *searching experiments* appear in expanded search within/out of focused factors; they can be modeled,

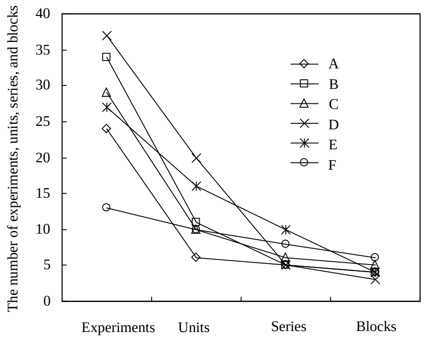

Fig. 5. Three levels of chunking by 6 subjects

from the viewpoint of the three levels of chunking, based on the patterns depicted in Figure 6 (b) through (d). In Figure 6 (b), after manipulating a single focused factor, subjects conduct expanded search out of the focused factor. We call this organization of searching behavior "second stage of chunking". In Figure 6 (c), although a subject manipulates two factors at once, the whole levels of the focused factors are not searched in the first unit; so expanded search within the focused factors appears. We call this organization process of behavior "third stage of chunking". In Figure 6 (d), both types of expanded search appear.

4.4 Compression Ratio of Chunking

We understand, through comparing Figure 5 and Figure 6, the behavior of Subject A represents the characteristics of FD experimental processes, whereas the behavior of Subject F represents TE search from the viewpoint of the three stages of chunking. The behavior of other four subjects represents the characteristics of *searching experiments* in which they organize their behavior on the second and third stages of chunking.

To clarify the discussions above, we define the compression ratio of chunking. The compression ratios on the first, second, and third stages of chunking are defined as the ratio of the number of Units to the total number of experiments, the ratio of the number of Series to the number of Units, and the ratio of the number of Blocks to the number of Series, respectively. As the compression ratio decreases, it means that higher compression is made. FD experimental behavior is structured only on the first stage of chunking on which higher compression is made; so the compression ratio is reduced from 1.0 whereas chunking on the second and third stages is not performed on which the compression ratios almost equal 1.0. On the other hand, in TE searching behavior the compression ratio on any stage of chunking nearly equals 1.0. The characteristics of behavior of *searching experiments* appear on the second and third stages of chunking on which the compression ratio is relatively reduced from 1.0.

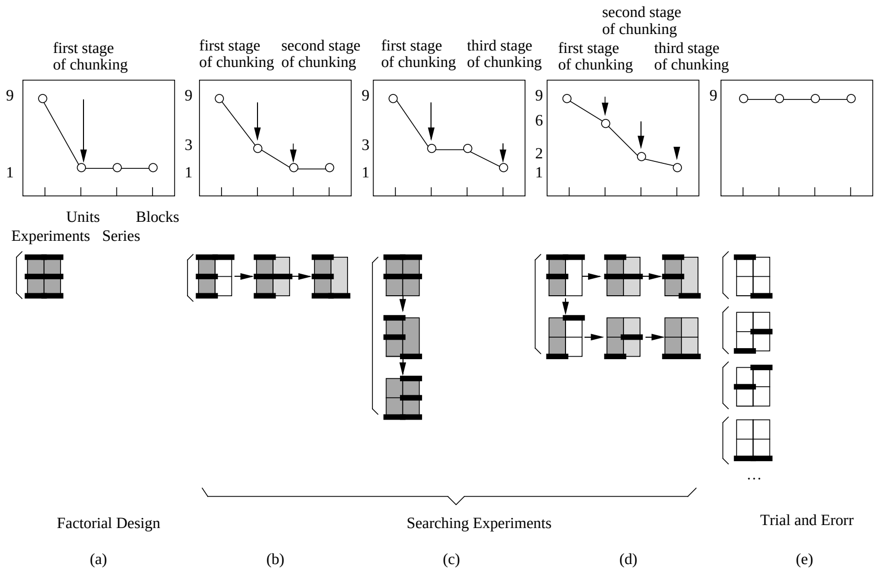

Fig. 6. Patterns of three stages of chunking

Figure 7 shows the compression ratios of each of the six subjects on the three stages of chunking. For example, let us consider an example behavior of Subject B dipicted in Figure 4. The compression ratios of Subject B on the three stages of chunking were .32, .45, and .80, because the number of experiments, Units, Series, and Blocks were, as seen in Figure 4, thirty-four, eleven, five, and four (also see Figure 5). Figure 7 indicates that the compression ratio of Subject A on the first stage is the smallest, and the compression ratios on the second and third stages nearly equal 1.0; so the behavior of Subject A reflects the characteristics of FD experimental processes. Subject F makes no chunks on any stage on which the compression ratio is relatively high; so the behavior of Subject F reflects the characteristics of TE search. In terms of other four subjects, chunking on the second or third stages, in addition to the first stage, is performed; so their behavior reflects the characteristics of *searching experiments*. Additionally, Figure 7 indicates that Subject D organizes his behavior on the second stage by expanded search out of focused factors because the compression ratio on this stage is smaller than that on the third stage. On the other hand Subject E organizes it on the third stage by expanded search within focused factors.

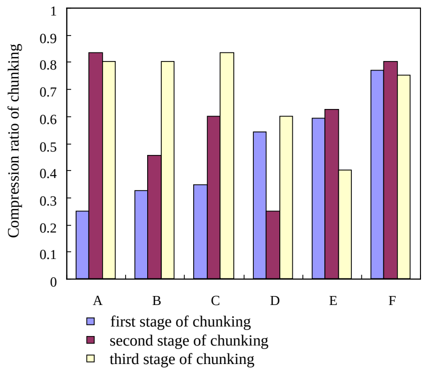

Fig. 7. Compression ratio of chunking

5 Searching Behavior and Performance of Discovery

5.1 Categorization of Final Solutions

Next we consider the relation between the characteristics of searching behavior described above and final solutions reached by each of the six subjects. In Table 3, searched factors related to the final solutions by the six subjects are indicated, each of which is classified from the following two viewpoints. First, the solutions are divided into two categories from the viewpoint of their generality. Solutions in one category refer to the factors that affect the system's performance while comparing several levels of the factors or mentioning to the effects of fluctuations of other factors. One example is "in terms of hypothesis formation strategies, the combination of the specific and general strategies produces the highest performance regardless of fluctuations of other factors." On the other hand, some subjects simply reported an individual level of searched factors that seem to decide the system's performance. One example is "in terms of hypothesis formation and testing strategies, when the former is the general strategy and the latter is the combination of the positive and negative testing strategies, the ratio of correct solutions reaches high." We call the former type of solutions general solutions whereas the latter specific solutions.

As the second viewpoint, the solutions in Table 3 are also classified from their validity. The correctness of each solution is decided based on both of knowledge on human discovery processes which has been obtained from cognitive psychological studies using the Wason's task [3] [4] and knowledge on regularities of our simulator's behavior identified in other our papers [5] [6]. We can divide the solutions of the six subjects into two categories from the two viewpoints mentioned above. One type of solutions is general and correct solutions whereas the other type is specific and incorrect solutions. Subject A, Subject B, and Subject C reached the former type of solutions, whereas Subject D, Subject E, and Subject F reached the latter type.

Table 3. Categorization of subjects' solutions

Subject	Factors	Generality	Validity
Subject A	AI-RH	General	Correct
Subject B	HF	General	Correct
Subject C	T T	General Specific	Correct Correct
Subject D	HT-HF	Specific	Incorrect
Subject E	HT-HF	Specific	Correct
Subject F	HT-HF	Specific	Incorrect

5.2 Factors Deciding Subjects' Performance

Now we move to discussions on the relation between the characteristics of subjects' behavior that were clarify in 4. and the solutions that each of the six subjects reached. Let us see Figure 7 again. The compression ratios of the subjects who reached general and correct solutions on the first stage of chunking are smaller than the ratios of those who reached specific and incorrect solutions. This indicates that the subjects who got correct and general solutions made higher compression on the first stage of chunking; those who got incorrect solutions could not. This insists that even though the characteristics of *searching experiments* appear on the second and third stages of chunking, chunking on the basic first stage is crucial for organizing their behavior.

Moreover Table 4 shows searched factors by each subject. The underlined factors indicate the factors related to the final solutions of each subject. The indexes, "o", "x", and "-", indicate systematically searched factors, randomly searched factors, and factors that were not searched, respectively. The systematic search means that the subjects searched whole levels of the factors or some representative levels, such as levels that have the highest or lowest values. Table 4 shows that two subjects, Subject B and Subject C, who reached correct solutions systematically searched two kinds of focused factors at once or conducted systematic search of other factors by expanded search out of focused factors. Subject A, even though he also reached correct solutions, did not conduct the expanded search. The reason is because the focused factors by Subject A, AI and RH, do not interact with other factors. On the other hand, every subject who reached incorrect solutions simply manipulated a single factor and could not conduct expanded search out of focused factors. Moreover some of them failed in systematic search of the focused factors.

Table 4. Searched factors and subjects' performance

Subject	Fucused Factors			Other Searched Factors	
Subject A	AI		o		-
	RH		o		-
	AI-RH		o		-
	HF		o		-
Subject B	T		x	HF	x
	T	HF	o	AI-RH	x
				AI-RH-TE	x
	HF	AI RH	x		
	T		o	HT	o
Subject C	T		o		-
	T	HT	o		-
	HT	HF	o	T	o
				TE	o
	T		o		-
	HF		o		-
Subject D	HT		o		-
	HT		o	T	x
	HF		o	T	o
Subject E	ambiguous		-		-
	AI-RH		x		-
	ambiguous		-		-
	HF		x		-
	AI		x	T	x
	TE	RH	o		-
Subject F	HT		o		-
	HF		o		-
	T		x		-
	HT		o	HF	x

6 Summary and Conclusions

In this paper, defining experimental processes that reflected the characteristics of both of FD experiments and TE search as *searching experiments*, we analyzed the ways of organizing behavior of *searching experiments* using SES. We also discussed the relation between the characteristics of the behavior and the performance of discovering targets. We understood, through the analysis of subjects' behavior on SES, that they organized their behavior on the three levels, Unit, Series, and Block. Chunking on the second and third stages, constructing Series and Blocks, reflected the characteristics of behavior of *searching experiments*.

In the latter part of this paper, we clarified that subjects who reached general and correct solutions effectively performed chunking on the first stage, which worked as the basis of organizing searching behavior, and systematically manipulated searched factors. One of our future works is to establish the ways for feedback of description of subjects' experimental processes based on SES and to discuss its educational effects.

References

1. Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research, Aldine Publishing Company.
2. Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
3. Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.
4. Klayman, J., & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 317-330.

5. Miwa, K., & Okada, T. (1996). Effective heuristics for hypothesis testing: an empirical discussion using a computer simulation method. *Journal of Japanese Society for Artificial Intelligence*, 11, 877-887.
6. Miwa, K. (1999). Collaborative Hypothesis Testing Process by Interacting Production Systems. *LNAI*, 1721, 56-67.
7. Okada, T. & Shimokido, T. (in press). The role of hypothesis formation in a community of psychology. In K. Crowley, C. D. Schunn, & T. Okada (Eds.) *Designing for Science: Implication from everyday, classroom, and professional settings.* Mahwah, NJ: Erlbaum.
8. Schunn, C., & Anderson, J. (1999). The Generality/Specificity of Expertise in Scientific Reasoning. *Cognitive Science*, 23, 337-370.
9. Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology.* 12, 129-140.