IEICE TRANS. INF. & SYST., VOL. E83-D, NO. 12 DECEMBER 2000

PAPER

A Study of Collaborative Discovery Processes Using a Cognitive Simulator

Kazuhisa MIWA†, *Regular Member*

SUMMARY We discuss human collaborative discovery processes using a production system model as a cognitive simulator. We have developed an interactive production system architecture to construct the simulator. Two production systems interactively find targets in which the only experimental results are shared; each does not know the hypothesis the other system has. Through this kind of interaction, we verify whether or not the performance of two systems interactively finding targets exceeds that of two systems independently finding targets. If we confirm the superiority of collaborative discovery, we approve of emergence by the interaction. The results are: (1) generally speaking collaboration does not produces the emergence defined above, and (2) as the different degree of hypothesis testing strategies that the two system use gets larger, the benefits of interaction gradually increases.
key words: *discovery, hypothesis testing, production system, collaboration, cognitive model*

1. Introduction

Various ways of acquiring empirical data in discovery processes are principally divided into two categories: experimentation and observation. In observation systems passively acquire data whereas in experiments systems selectively gather data based on their hypotheses. In natural sciences, such as physics, chemistry, and biology, experimentation plays an essential role. Therefore both hypothesis testing and hypothesis generation processes are important research topics.

In terms of the studies of hypothesis testing, philosophers have tried to find normative principles of hypothesis testing (i.e., how people should think), whereas psychologists have tried to clarify the practical usage of how humans actually conduct hypothesis testing [2]. In psychological research, it has been known that there are several seemingly irrational cognitive biases in human hypothesis testing.

Many historically important discoveries have been made not only by a single genius but also through collaborative activities of several persons. This fact indicates that interaction between several agents (i.e. humans, computers, and systems) sometimes produces positive effects. Our question is whether or not collaboration by humans who have the cognitive biases mentioned above produces these effects.

It is important to define a concept of “emergence” when we consider collaborative problem solving. We intuitively understand that the performance of two persons solving a problem is usually better than that of a single person solving it. But we do not approve of this advantage as the appearance of emergence.

To discuss the criterion of introducing emergence, we consider the following three situations (see Fig. 1).

Single situations: A problem solver tries to discover a target. If the final hypothesis obtained through conducting his/her experiments is identical to the target, then we confirm that he/she finds the target.

Independent situations: Two problem solvers independently conduct experiments and form hypotheses without interaction. If at least one of them reaches a correct final hypothesis that is identical to a target, then we confirm that they independently find the target.

Collaborative situations: Two problem solvers interact with each other. That is, each person alternatively conducts experiments, and each experimental result is given to both persons. Half of their acquired data comes from his/her self-experiments and the other half from the other person’s experiments. If at least one of them reaches a final correct hypothesis, we confirm that they cooperatively find the target.

To estimate whether or not emergence appears, we do not compare the performance in the single situations and the performance in the collaborative situations. Otherwise, when the performance in the collaborative situations exceeds the performance in the independent situations, we confirm the emergence by interaction between the two problem solvers.

In this paper, on the basis of the above viewpoints, we discuss the possibility of emergence in human collaborative discovery processes using a computational model that is constructed as a production system model. Another approach for discussing the effects of interaction between humans is to use psychological experiments. The computational method used in this paper and the experimental method are two primary approaches for investigating human discovery processes in the field of cognitive science.

Two methods are strongly related; when a computational model predicts empirical results, through confirming the results by psychological experiments assumptions in the model are verified. In this paper we use the former approach to utilize advantages of the computational approach. One of the advantages is that we can test many experimental conditions by modifying the models. It is often difficult to do so in psychological experiments because the number of subjects participated in the experiments are restricted. Additionally the detailed control of experimental conditions by instructing subjects, which will be conducted in our simulations, is some-

Manuscript received February 18, 2000.
Manuscript revised July 10, 2000.
† The author is with the Graduate School of Human Informatics, Nagoya University, Nagoya-shi, 464-8601 Japan.

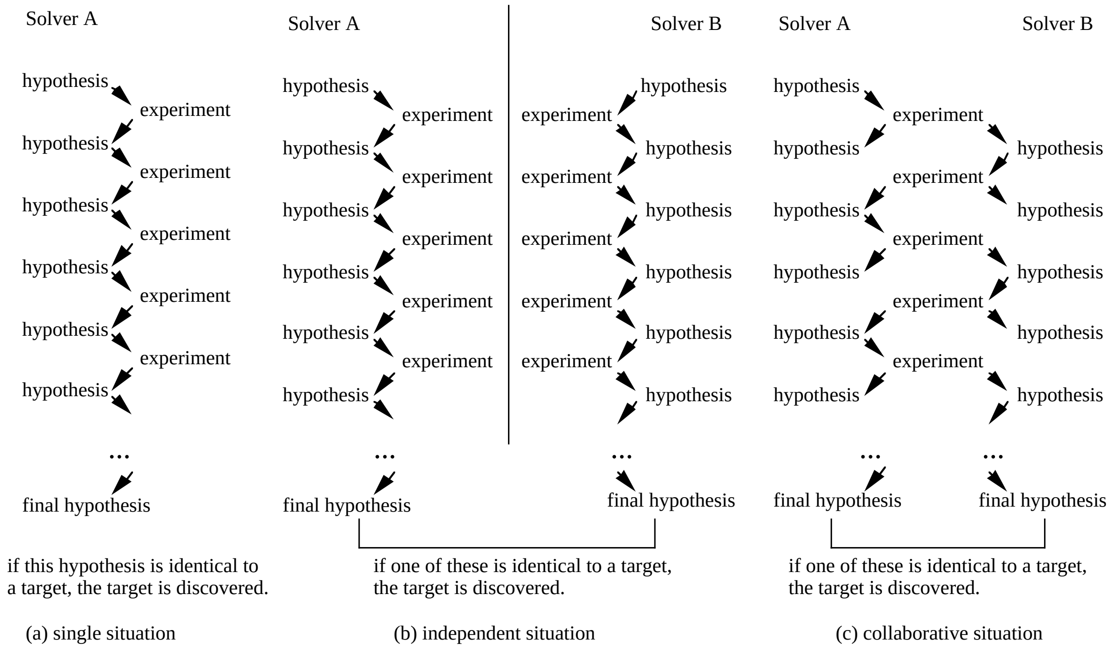


Fig. 1 Single, independent and collaborative situations.

times impossible. Of course we understand that the results of our computer simulations in this paper should be supported by empirical data in our future work.

2. Research Background

2.1 Human Hypothesis Testing

One style of experimental psychological studies on human hypothesis testing is analyzing hypothesis testing processes of human subjects in a psychological experimental room using a relatively simple discovery task [4]. One of the most famous tasks used in those experiments is the Wason's 2-4-6 task [14], [17]. In this paper, we also use this task.

The most basic distinction of hypothesis testing strategies is a positive test (Ptest) and a negative test (Ntest). Ptest is hypothesis testing that generates a positive instance for the hypothesis whereas Ntest generates a negative instance. Many empirical studies have indicated that humans have a positive test bias in which they tend to test their hypothesis by using a positive instance for the hypothesis. On the other hand, philosophers have stressed the importance of hypothesis falsification [16]. It seems that there is a contradiction between the empirical data brought by psychological experiments and the normative principle of hypothesis testing by philosophers.

Klayman and Ha indicated on the basis of their sophisticated analysis that the possibility of hypothesis disconfirmation and the usage of positive testing were essentially independent [6], [7]. Their conclusions were that when targets to be found were general (the ratio of target instances to all instances is high), negative testing was effective for disconfirming hypotheses; on the other hand, to find specific targets (the ratio of target instances to all instances is low), positive testing was more effective than negative testing for disconfirmation. In ordinary contexts actual targets are usually specific, so the positive test bias is not irrational for a hypothesis testing strategy.

2.2 Human Collaborative Discovery

There have been some proceeding psychological studies in which human collaborative discovery processes were analyzed using similar tasks such as the 2-4-6 task. In most of these studies, the performance of a single person and the performance of a group consisting of two to four persons were simply compared. The definition of emergence in this paper (i.e., independent and collaborative solving by two persons was compared) was not used in these studies. So in the discussions below, on the basis of the following procedure, we calculated the performance in the independent condition from the performance of a single person. That is, when independent n persons whose performance was m ($0 < m < 1$) solved a problem, the probability of at least one of them reaching the solution is $1 - (1 - m)^n$. We used this score as the performance in the independent solution.

Then we analyzed the previous studies based on the procedure above. We found that empirical knowledge on the possibility of emergence is not consistent in these studies. For example, Lauglin and her colleagues compared the performance of a single person and that of a group of four persons with a discovery task using playing cards [8]. According to the score defined above, the performance of group collaborative problem solving does not exceed the performance of independent solving by four persons. On the other hand, Okada and Simon analyzed discovery processes of a single subject and those of two collaborative subjects using a microworld called a simulated molecular genetics laboratory [15]. Consequently, they confirmed the appearance of emergence by the interaction of two collaborative subjects. Some other studies also support the experimental results indicating the predominance of group problem solving to independent problem solving [1], [3].

2.3 Collaboration when Experimental Data are Only Shared

It is important to note the difference between the situations of the proceeding studies and the situation of our simulations in this paper. In our simulations, two production systems only share experimental results. Each system does not know which hypothesis the other system has. That is, the experimental space is shared but the hypothesis space is not shared [5]. This situation in the simulations corresponds to the situation in which two persons exchange only mutual experimental results without conversation. However in the proceeding psychological studies, conversational interaction between subjects was not restricted.

The effects of collaboration seem to appear in several stages of interaction: (1) interaction of hypothesis formation, (2) interaction of experimental verification, and (3) integration of the two kinds of interaction. In this paper, we focus on the second type of interaction above. Other types of interaction are discussed in our other papers [12].

The point is that the second type of collaboration dealt with in this paper needs neither additional working memory capacity nor new production rules. Two systems only exchange mutual experimental results. When the first and third types of interaction are allowed, each of two problem solvers needs additional abilities for processing information introduced from the other solver: for example, an ability for forming a different hypothesis while referring another hypothesis of the other solver, or an ability for designing experiments by combining two kinds of hypotheses. However, in the second type of interaction, these additional abilities are not needed. Two problem solvers only exchange mutual experimental results.

Our question is whether or not emergence occurs in this type of collaboration. If we observe emergence in this collaboration, we find an important function of collaborative problem solving because we can obtain some benefits from interaction without any additional cognitive abilities.

Now I'd like to summarize the objectives in this paper.
(1) We discuss the possibility of emergence in the situations in which only experimental results are shared.
(2) Humans have a positive test bias in hypothesis testing. We discuss the relation between the combinations of these biases (generally speaking, the differences of subjects' hypothesis testing strategies) and the benefits of interaction.

3. Cognitive Simulator

We use our computational model as a cognitive simulator. In this section, we discuss basic concepts of a cognitive simulator and the specifications of the simulator.

3.1 Experimental Design

The objective of using a cognitive simulator is to conduct psychological experiments by simulating human problem solving on a computer. The framework of designing computer simulations corresponds to the experimental design that has been established in experimental psychology.

In planning an experimental design, the experimental factors are controlled by the experimenter's instructions to subjects and experimental situations; each factor is assigned to each group of subjects or each trial of the experiments that one subject repeatedly performs. The relation between the controlled factors (independent variables) and the performance (a dependent variable) is discussed. Statistical methods such as ANOVA are widely used to determine the relation of the independent and dependent variables. In using a cognitive simulator, controlling factors is conducted by managing the parameters of a model which decide the model's behavior.

3.2 Interactive Production System

We have developed an interactive production system architecture for constructing the cognitive simulator and simulating collaborative discovery processes. 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 Fig. 2). 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.

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.

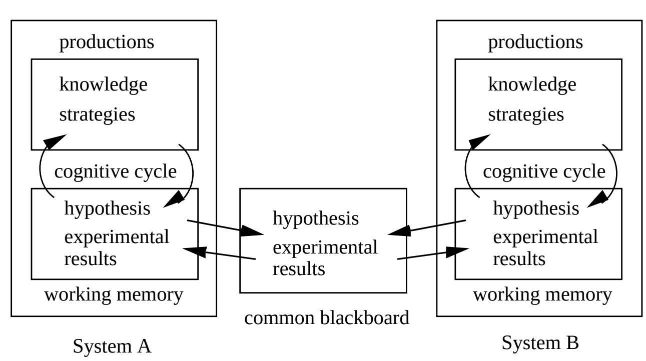


Fig. 2 The architecture of the interactive production system.

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

Basically the model searches the hypothesis space randomly in order to generate a hypothesis, though 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.

You can see the detailed specifications of this model in Miwa and Okada, 1996 [11].

3.3 Wason's 2-4-6 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.

3.4 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.

4. Psychological Validity of the Simulator

4.1 Laughlin's Experiments (1997)

To what degree does the simulator used in this paper correctly reflect actual human hypothesis testing processes? Recently, based on the rule induction paradigm [9] as a style of cognitive psychological experiments, Laughlin and her colleagues analyzed collaborative discovery processes by a group consisting of four members, each of whom used a different hypothesis testing strategy [10]. The task used in the experiments was another a discovery task called New Elusis. Their interests were: (1) the relation between the combinations of hypothesis testing strategies used by four group members and their total performance, and (2) the ratio of acquiring target instances from subjects' experiments. We attempted to verify whether our cognitive simulator correctly predicts Laughlin's experimental results.

4.2 Conditions of Preliminary Simulations

There are several differences between Laughlin's experimental situation and the situation in which our cognitive simulator is used. To reduce the differences, we tuned up the specifications of our simulator as follows:
(1) In Laughlin's experiments the condition in which all four members were forced to conduct Ptest was called the PPPP condition. They set up six conditions: the PPPP condition, the PPPN condition (three members conducted Ptest, and the other member conducted Ntest), the PPNN condition, the PNNN condition, the NNNN condition, and the control condition (each member selected hypothesis testing as he/she liked). In our simulations, two production systems (not four members) interacted. So we selected three conditions, the PPPP, PPNN, NNNN conditions, from the six conditions of Laughlin's experiments and compared the Ptest vs. Ptest condition, the Ptest vs. Ntest condition, and the Ntest vs. Ntest condition in our simulations to the three conditions, in which the Ptest condition means that a system always uses Ptest in experiments, and the Ntest condition means that a system always uses Ntest.
(2) We used 35 targets in our main simulations (see Sect. 5). The patterns of hypothesis confirmation and disconfirmation were strongly influenced by the degree of generality of a found target (see Sect. 2.1). So in the preliminary simulations, we selected 8 targets from the 35 targets whose generality is similar to that of the targets used in Laughlin's experiments.
(3) In Laughlin's experiments, four members shared their hypotheses in addition to the result of each experiment, then selected one of the hypotheses as a group hypothesis through their conversation. In our simulations, on the other hand, only the results of experiments were shared. To reduce the difference between the two experiments, we utilized the following procedures. In the preliminary simulations one of the two hypotheses that two production systems independently formed was randomly selected as a group hypothesis, and then an experiment was conducted on the basis of the randomly selected group hypothesis.
(4) In the main simulations (see Sect. 5.), when each hypothesis of two systems is confirmed by continuous four experiments, the systems terminated their experiments and confirmed the final hypothesis as the solution. So the total number of experiments varied in every simulation. On the other hand, in Laughlin's experiments the number of experiments was fixed; eleven experiments per subject (a total of 44 experiments by four members) were permitted until the members decided the final hypothesis. So in the preliminary simulations the number of experiments is fixed; that is, when a total of 14 experiments is done, the systems terminate their experiments.

4.3 Results of Preliminary Simulations

Figure 3 shows the comparison of the results of our computer simulations (the lower side) and the result of Laughlin's experiments (the upper side). The left side of the figure indicates the comparison of the ratio of correct hypotheses matched to the targets throughout the discovery processes, where the horizontal axis indicates the PPPP (Ptest vs. Ptest), PPNN (Ptest vs. Ntest), NNNN (Ntest vs. Ntest) conditions.

The right side indicates the ratio of target instances obtained from experiments, where at the PPNN (Ptest vs. Ntest) condition, the ratio by one subject (system) conducting Ptest and the ratio by another subject conducting Ntest are separately indicated.

Figure 3 shows the pattern of the ratio of correct hypotheses in our simulations is similar to the results of Laughlin's experiment. However, the ratio of target instances in both cases differs to some degree, especially in the NNNN (Ntest vs. Ntest) condition.

Laughlin et al. indicated that subjects in the NNNN condition used the strategic hypothesis testing strategy, a strategy which was not used in other conditions. In the NNNN condition, subjects rarely observed target instances. To observe target instances, they formed a secondary strategic hypothesis and then generated a negative instance for the secondary hypothesis; this negative instance was simultaneously a positive instance for the primary hypothesis. Our simulator cannot produce this phenomenon because of the absence of

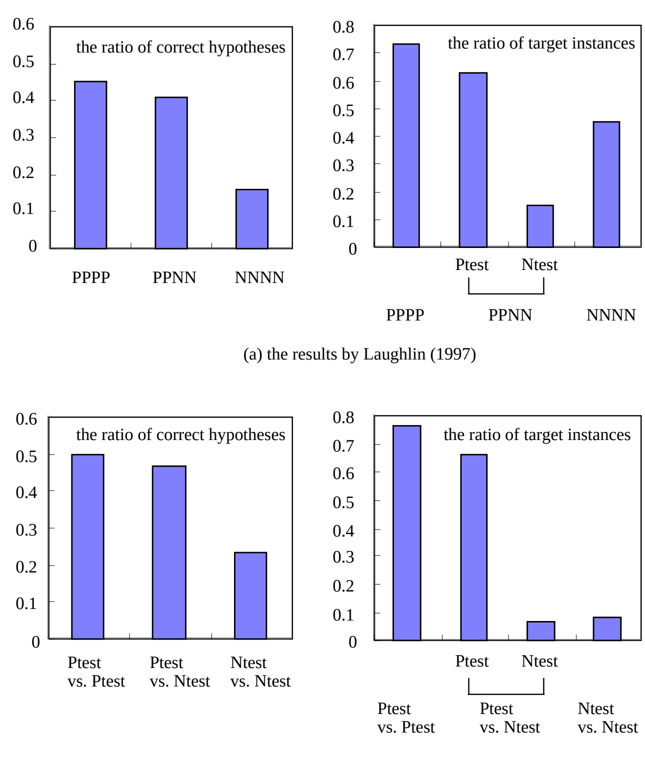


Fig. 3 The results of the preliminary simulations and the comparison between the results and Laughlin's experiments.

the mechanism for conducting strategic hypothesis testing. However the results of our simulations replicated the pattern of confirmation and disconfirmation of hypotheses by human subjects well in terms of simple combinations of Ptest and Ntest.

5. Results of Computer Simulations

Based on the verification of our cognitive simulator in 4., we move to the main simulations.

5.1 Conditions of Main Simulations

In the following main simulations, we let our simulator find 35 targets in each simulated condition. Table 2 shows the targets used in the simulations.

First, we conducted 30 simulations in finding each target and obtained the ratio of correct solution for each target. Then we calculated the average ratio of correct solutions for finding the 35 targets. We used the average ratio as a score of performance.

5.2 Single, Independence, and Collaboration

In this subsection, first we compare the performance of a single system finding the targets and that of dual systems finding the targets.

Second, we also test whether or not emergence defined in 1. appears by comparing the performance of two collaborative systems finding the targets and the performance of two independent systems finding the targets. To do so, we conduct simulations based on the following 4 * 6 experimental design. The contents in () indicate the levels of each factor.

Table 2 Thirty five targets used in the main simulations.

dimensions	targets
Order	ascending numbers, equal or ascending numbers
Interval	the interval is 2, the interval is same
Even-ODD	continuous evens, three evens
Range of desits	single digits, positive digits
Certain digit	the first number is even, the second is even, the third is even, the first is 2, the second is 4, the third is 6
Multiples	multiples of 2
Divisors	divisors of12, divisors of 24
Sum	the sum is even, the sum is a double digit, the sum is a positive number, the sum is 12, the sum is a multiple of 12, the sum is a multiple of 6, the sum is a multiple of 4, the sum is a multiple of 3, the sum is a multiple of 2
Product	the product is even, the product is a double digit, the product is a positive number, the product is 48
Different	different three numbers
Mathematical relationship	the third = the first + the second, the third = the first * the second - 2, the third = the first * 3 and the second = the first * 2, the third = the first * n and the second = the first * m

Design

Factor A: single/independent/collaborative conditions (Single A, Single B, Independent, Collaborative)

Factor B: combinations of hypothesis testing strategies (Ptest vs. Ptest, Mtest vs. Mtest, Ntest vs. Ntest, Ptest vs. Mtest, Mtest vs. Ntest, Ptest vs. Ntest)

The first two levels of factor A (i.e., Single A and Single B) correspond to the cases in which a single system finds the targets. One system indicating the lower performance is assigned to Single A, and the other system indicating the higher performance to Single B. The latter two levels, Independent and Collaborative, correspond to the cases in which dual systems independently and collaboratively find the targets.

In the computer simulations, we also consider the combinations of hypothesis testing strategies that two systems use. We utilize three kinds of strategies: Ptest, Ntest, and Mtest strategies. The Ptest strategy always uses Ptest in experiments, the Ntest strategy always uses Ntest, and the Mtest (Medium test) strategy uses Ptest in the half of experiments and uses Ntest in the other half. The combinations of these three strategies are 6 cases ((a) Ptest vs. Ptest, (b) Ntest vs. Ntest, (c) Mtest vs. Mtest, (d) Ptest vs. Mtest, (e) Ntest vs. Mtest, and (f) Ptest vs. Ntest).

Results

The results of the computer simulations are indicated in Fig. 4. The vertical axis indicates the average ratio of finding the 35 targets, and the horizontal axis indicates each level of the two factors. Figure 4 shows the following tendencies.

First, in every combination of hypothesis testing strategies, the performance of the dual systems exceeded the performance of a single system. In terms of the independent and collaborative solutions, the performance in both cases did not differ much, or the performance in the former case sometimes exceeded in the latter case.

Next, for a more detailed analysis, we conducted an

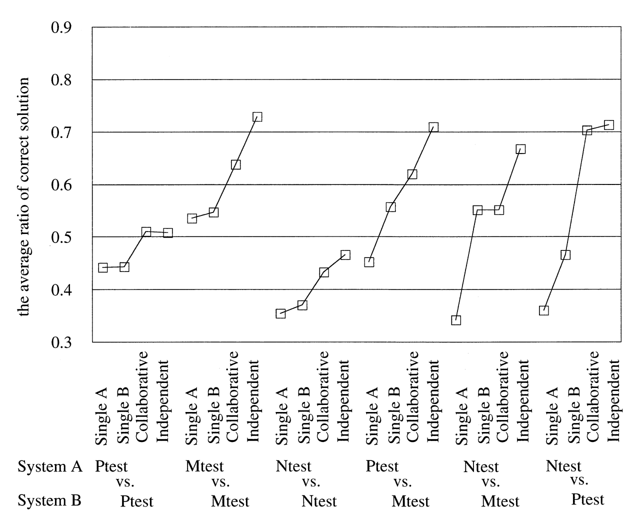


Fig. 4 Comparison of the performance in the single, independent, and collaborative situations.

ANOVA. The interaction between Factor A and Factor B was significant. So we analyzed the main effect of Factor A individually in each of the 6 levels of Factor B. As a result, we confirmed the main effect of Factor A in every level of Factor B. Moreover, based on the LSD (Least Significant Difference) analysis, we confirmed that the performance of collaborative and independent discovery in the Ptest vs. Ptest, Ntest vs. Ntest, and Ptest vs. Ntest conditions did not differ, whereas the performance of independent discovery exceeded that of collaborative discovery in the Mtest vs. Mtest, Ptest vs. Mtest, and Mtest vs. Ntest conditions.

The experimental results above show that the collaborative discovery did not produce emergence, but sometimes indicated a lower performance to the performance of the independent discovery.

5.3 Robustness of the Results

To what degree can the results above be generalized? In other words, are these results above independent of other factors? To confirm the robustness of the experimental results in Sect. 5.2, what we should indicate is that there is no interaction between Factor A and Factor C in each level of Factor B, where Factor C is an assumed third factor for checking the robustness. The absence of interaction between Factor A and Factor C indicates that the results in Sect. 5.2 are not influenced by this third factor, Factor C.

In the following sections, we will compare only the performance of independent and collaborative discovery and omit the comparison of the performance of a single system and dual systems. Thus, we conduct the analysis above, but narrow the scope of discussions to the two levels of Factor A, Independent and Collaborative.

[1] the condition of terminating experiments
First, we consider the condition of terminating experiments as Factor C. This condition decides when systems terminate their experiments. In the simulations in Sect. 5.2, if four continuous confirmations were obtained, systems terminated experiments and confirmed the final hypothesis formed at that time as the solution. We considered other options in terms of this condition; that is, three and five continuous confirmations may be enough for terminating experiments. We predicted that as this number increases, the systems become more deliberate for the termination of exploring a target and the performance increases.

To analyze this, we conducted the following simulations based on the 2 * 6 * 3 experimental design.

Design

Factor A: independent/collaborative conditions (Independent, Collaborative)

Factor B: combinations of hypothesis testing strategies (6 levels (see Sect. 5.2))

Factor C: the condition of terminating experiments (3, 4, and 5 continuous confirmations)

Results

Figure 5 shows the results of the simulations. According to the ANOVA, there is no interaction between Factor A and Factor C at all levels of Factor B. The main effect of Factor C is significant. These results support the conclusions that the comparison of the performance of independent and collaborative solutions obtained in Sect. 5.2 are not influenced by the condition with procedures terminating experiments.

[2] working memory capacity
Next we examined the robustness against the fluctuation of a working memory ability. To check this effect, we conducted further simulations based on the following 2 * 6 * 4 experimental design.

Design

Factor A: independent/collaborative conditions (Independent, Collaborative)

Factor B: combinations of hypothesis testing strategies (6 levels (see Sect. 5.2))

Factor C: working memory capacity (3, 4, 6, and whole instances)

Each level of Factor C indicates the number of the instances that can be simultaneously activated in working memory. For example, if the working memory capacity is 3 but ten instances are obtained, the systems form their hypothesis based on only three instances that are randomly selected from the ten instances. In the simulations in Sect. 5.2, whole instances can be activated when forming a hypothesis. We predict that as the number of activated instances decreases, the performance decreases.

Results

Figure 6 indicates the results of the simulations. According to the ANOVA, the interaction between Factor A and Factor C reaches significance in two levels of Factor B, Ntest vs.

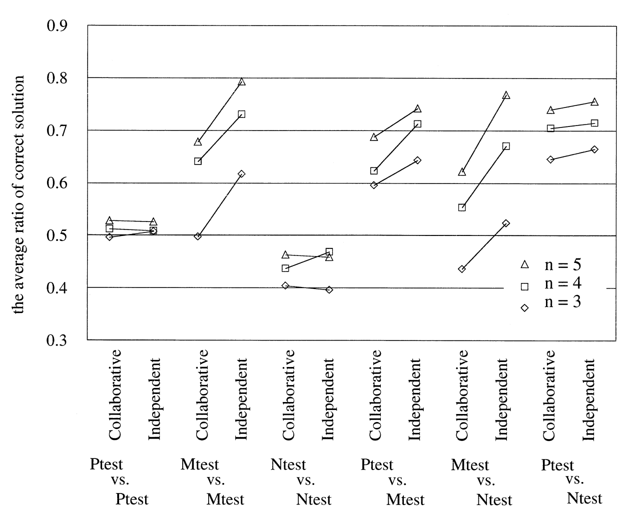


Fig. 5 Comparison of the performance while varying the condition for the termination of experiments.

Ntest and Ptest vs. Mtest. In the other four levels, the interaction between Factor A and Factor C was not significant.

In Fig. 6, the pattern of the performance in the level with 3 activated instances was different than the other three levels (i.e., 4, 6 and whole instances). When the homogeneous three levels (4, 6, and whole instances) are examined while excluding the different level with 3 instances, the interaction between Factor A and Factor C disappears at every level of Factor B.

The above discussions show that the behavior of this simulator is robust against the fluctuation of the working memory capacity if the number of simultaneously activated instances is not below 4.

5.4 Difference of Hypothesis Testing Strategies

We move to further analysis of the relation between the benefits of interaction and the different degree of hypothesis testing strategies that two systems use because we confirmed a certain amount of robustness of our simulator through the results and discussion above.

To do so, in the following simulations, we control the degree of difference in hypothesis testing strategies of the two systems while keeping the total ratio of conducting Ptest by the two systems at 50%.

Design

Factor A: independent/collaborative conditions (Independent, Collaborative)

Factor B: combinations of hypothesis testing strategies (50% vs. 50%, 62% vs. 38%, 75% vs. 25%, 87% vs. 13%, 100% vs. 0%)

In terms of Factor B, the percentage of each level indicates the probability of each system conducting Ptest in its experiments. So the 50% vs. 50% condition and the 100% vs. 0% condition correspond to the Mtest vs. Mtest condition and the Ptest vs. Ntest condition in Sects. 5.2 and 5.3, respectively.

Results

Figure 7 shows the results of the simulations. Figure 7 indicates that as the different degrees of hypothesis testing strategies gets larger, the performance in the condition of collaborative discovery gradually increases whereas the performance in independent discovery remains almost constant.

According to the ANOVA, the interaction between Factor A and Factor B is significant. The main effect of Factor A in the 50% vs. 50%, 62% vs. 38%, and 75% vs. 25% conditions is significant whereas it is not significant in the 87% vs. 14% and 100% vs. 0 % conditions.

The analysis above indicates the benefits of interaction increases as the different degree of hypothesis testing strategies increases.

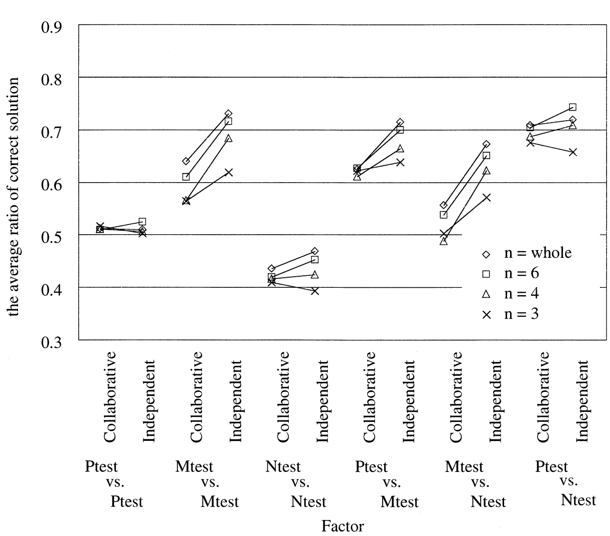


Fig. 6 Comparison of the performance while varying the working memory capacity.

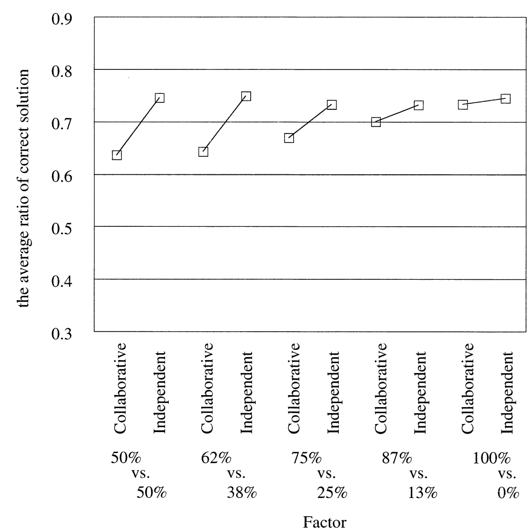


Fig. 7 The relation between the different degree of hypothesis testing strategies of the two systems and their total performance.

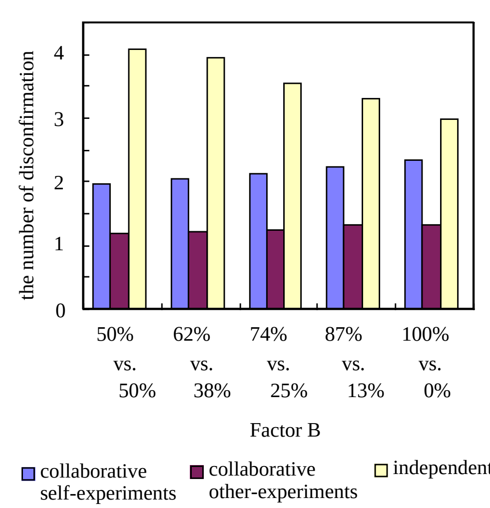
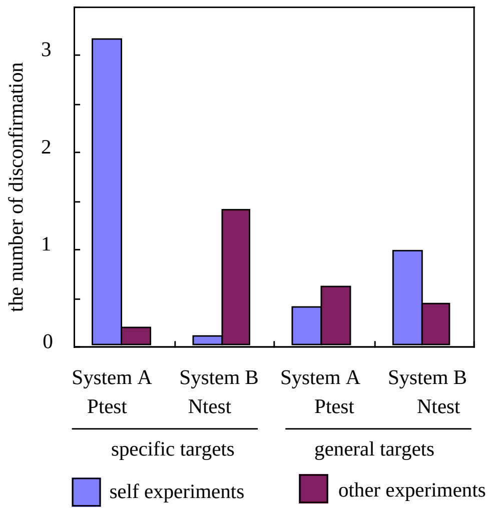


Fig. 8 The relation between the different degree of hypothesis testing strategies and the number of disconfirmation



(a) Ptest (100%) vs. Ntest (0%)   (b) Mtest (50%) vs. Mtest (50%)

Fig. 9 The pattern of disconfirmation in the interactive situations.

5.5 Improvement of Disconfirmation Ability by the Combination of Different Strategies

Where does the improvement of the performance by applying different hypothesis testing strategies come from? As we mentioned before, the performance of systems strongly depends on the ability of the systems disconfirming hypotheses. Figure 8 indicates the relation between the different degree of the strategies and the average number of disconfirmations that two systems introduce until the experiments are terminated. In the case of collaborative discovery, disconfirmation was introduced by both of the self-conducted experiments and the experiments that the other system conducts; thus, both kinds of disconfirmation are separately indicated. In the case of collaborative solutions, the ability of disconfirmation increased as the different degree of the strategies became larger, whereas in the case of independent solutions, the ability decreased. How does this kind of improvement of disconfirmation ability in the collaborative situation appear? As we mentioned before, the ability of disconfirmation of Ptest and Ntest is different, depending on the generality of the targets that the systems try to find.

So we divided the 35 targets into 18 specific targets and 17 general targets, and investigated the occurrence pattern of disconfirmation in both cases in the collaborative problem solving situations (see Fig. 9). When each system used the Mtest strategy (the 50% vs. 50% condition), the occurrence pattern of disconfirmation of the two systems was almost the same because both systems used the same strategy. However, in the 0% vs.100% condition, there was a big difference.

When the problem was to find the specific targets, System B, which used the Ntest strategy, disconfirmed its hypotheses more by the other-conducted experiments than by self-conducted experiments. In contrast, System A, which used the Ptest strategy, did more by self-conducted experiments. When finding the general targets, this tendency was reversed.

We pointed out that the Ptest strategy promoted disconfirmation when finding specific targets whereas using Ntest prevented the systems from finding specific targets (see Sect. 2.1). So the results above show that a system that uses a disadvantageous strategy disconfirms its hypotheses by experiments conducted by the other system that uses an advantageous strategy. This result can be understood through the following interpretation: for finding various targets whose nature is different (e.g., specific and general), the lack of abilities of one system that used a disadvantageous strategy is supplemented by the other system that used an advantageous strategy. This kind of complementary iteration produced an additional ability for disconfirming hypotheses, and accounted for the superiority of the two systems using different strategies.

Some parts of these results could be theoretically predicted based on the Klayman's analysis shown in Sect. 2.1. However we believe that confirming the theoretical prediction through computer simulations is essentially important. Additionally, in the studies of collaborative problem solving, the importance of two problem solvers having different viewpoints has been indicated. Our results empirically support this idea.

6. Conclusions and Future Works

In this paper, we investigated the processes of two systems collaboratively finding targets using a cognitive simulator.

The results are summarized as follows.

(1) Generally speaking, collaboration by only sharing the experimental space, that is, only experimental results are exchanged, does not produce emergence. Solving a problem independently is more profitable than solving it collaboratively.

(2) As the different degree of hypothesis testing strategies becomes larger, the benefit of collaboration increases.

(3) This benefit results from the improvement of the systems' disconfirmation ability. The collaboration between the two systems that use different hypothesis testing strategies produces the complimentary interaction in which a decline of the ability by using a disadvantageous strategy is supplemented by the other system using an advantageous strategy. This complementary interaction improves the total performance of the two collaborative systems.

Even when the difference of two systems' testing strategies reaches the maximum, the performance of two interacting systems did not exceed that of two independently solving systems and we could not confirm the appearance of emergence. Our next question is in which cases does emergence appear? We can consider the following two possibilities.

First, we can consider the cases in which we let the difference of strategies of two systems be much bigger than in the current situations demonstrated in this paper, controlling hypothesis formation strategies in addition to hypothesis testing strategies. Second, we consider the cases in which we increase the stages of the shared cognitive space; for example, the hypothesis space is also shared in addition to the experimental space. In this case, each system knows the hypothesis that the other system has: this kind of interaction was not permitted in the current simulations. We have already begun to examine these next steps and obtained some interesting results [12], [13].

References

[1] E. Freedman, "Scientific induction: Individual versus group processes and multiple hypotheses," Proc. 14th Annual Meeting of Cognitive Science Society, Lawrence Erlbaum Associates Ltd, 1992.

[2] R.N. Giere, ed., Cognitive Models of Science, the University of Minnesota Press, 1992.

[3] M. Gorman, M. Gorman, M. Latta, and G. Cunningham, "How disconfirmatory, confirmatory and combined strategies affect group problem solving," British Journal of Psychology, vol.75, pp.65-79, 1984.

[4] M.E. Gorman, Simulating science: Heuristics, mental models, and technoscientific thinking, Indiana University Press, 1992.

[5] D. Klahr and K. Dunbar, "Dual space search during scientific reasoning," Cognitive Psychology, vol.12, pp.1-55, 1988.

[6] J. Klayman and Y.-W. Ha, "Confirmation, disconfirmation, and information in hypothesis testing," Psychological Review, vol.94, pp.211-228, 1987.

[7] J. Klayman and Y.-W. Ha, "Hypothesis testing in rule discovery: Strategy, structure, and content," J. Experimental Psychology: Learning, Memory, and Cognition, vol.15, pp.596-604, 1989.

[8] P.R. Laughlin & C.G. Futoran, "Collective induction: Social combination and sequential transition," J. Personality and Social Psychology, vol.48, pp.608-613, 1985.

[9] P.R. Laughlin, "Group decision making and collective induction," in eds. E. Witte and J. Davis, Understanding Group Behavior, Lawrence Erlbaum Associates, 1996.

[10] P.R. Laughlin, V.J. Magley, and E.I. Shupe, "Positive and negative hypothesis testing by cooperative groups," Organizational Behavior and Human Decision Processes, vol.69, pp.265-275, 1997.

[11] K. Miwa and T.Okada, "Effective heuristics for hypothesis testing: An empirical discussion using a computer simulation method," J. Japanese Society for Artificial Intelligence, pp.877-887, 1996

[12] K. Miwa, "Collaborative discovery processes by interactive production systems: Degree of sharing cognitive space and possibility of emergence," Proc. 2nd International Conference of Cognitive Science, pp.343-348, 1999.

[13] K. Miwa, "Collaborative hypothesis testing process by interacting production systems," LNAI, vol.1721, pp.56-67, 1999.

[14] S. Newstead and J. Evans, eds., Perspectives on Thinking and Reasoning, Lawrence Erlbaum Associates Ltd., UK, 1995.

[15] T. Okada and H. Simon, "Collaborative discovery in a scientific domain," Cognitive Science, vol.21, pp.109-146, 1997.

[16] K. Popper, The logic of scientific discovery, Harper & Row, 1959.

[17] P.C. Wason, "On the failure of eliminate hypothesis in a conceptual task," Quarterly Journal of Experimental Psychology, vol.12, pp.129-140, 1960.

Kazuhisa Miwa 1989, Dr. of Eng. Department of Information Engineering, Nagoya University. 1989-1993, Assistant Professor of Education Center for Information Processing, Nagoya University. 1991-1992, Visiting Assistant Professor of Department of Psychology, Carnegie Mellon University. 1993-, Associate Professor of Graduate School of Human Informatics, Nagoya University. Research Interests: Computational Models for Human Problem Solving, Learning, Scientific Discovery. Affiliations: Japanese Society for Artificial Intelligence; The Japanese Cognitive Science Society.