Emergence of Effects of Collaboration in a Simple Discovery Task

Kazuhisa Miwa (miwa@cog.human.nagoya-u.ac.jp)
Graduate School of Human Informatics, Nagoya University
Nagoya, 464-8601 JAPAN

Abstract

We discuss the effects of collaboratively finding a target in a simple discovery task, using the Wason's 2-4-6 task. We control the following two factors: hypothesis testing strategies that participants use, and the nature of targets that they find. First, we propose, through computer simulations, a hypothesis on a situation in which the effects emerge. Then we verify the hypothesis by psychological experiments. Last, we generalize, through theoretical analysis, the findings obtained by the two empirical approaches above. As the result, it has been concluded that the effects of collaboration emerge in the following situations: (1) two participants repeatedly conduct a positive test for finding a general target, and (2) each of them maintains a different hypothesis.

Introduction

In psychological studies on scientific discovery, relatively simple tasks, such as the Wason's 2-4-6 task and New Elusis, have been so far used (Gorman, 1992; Newstead & Evans, 1995). In recent days, using those tasks, the effects of collaboration have been empirically discussed when several participants collaboratively find a target.

We usually think that working together provides positive effects. However, those empirical results obtained in the psychological studies above do not necessarily support the intuitive prediction.

In these studies, the performances (the proportions of correct findings) in the single condition in which a single subject performs the task and those in the collaborative condition in which a group of n subjects collaboratively performs the task are compared. In this comparison, even when the latter performance exceeds the former, the advantage may be introduced not by the interaction among the subjects, but simply by the quantity of the subjects. That is, in the latter case n solutions (final hypotheses) by the n subjects are considered, and the probability of that at least one of them is accidentally identical to the target is much higher than that in the former case. So we should consider the independent condition in which n participants independently perform the task without interaction. The performance in the independent condition can be theoretically calculated from the performance in the single condition. That is, the probability of that at least one of n subjects reaches the solution is $1 - (1 - p)^n$ where the probability of each subject's finding the correct target is p $(0 < p < 1)$. We utilize this score as the performance in the independent condition.

Table 1 reviews the comparison of the performances in the single, independent, and collaborative conditions in the preceding studies (Freedman, 1992; Laughlin & Futoran, 1985; Laughlin & McGlynn, 1986; Laughlin, VanderStoep, & Hollingshead, 1991; Laughlin, Bonner, & Altermatt, 1998; Okada & Simon, 1997). Table 1 shows that the performance in the collaborative condition cannot exceed that in the independent condition in almost all cases.

In this study, we will discuss states in which the effects of collaboration emerge based on the results above. As an approach, first we will propose a hypothesis on when the effects of collaboration appear by computer simulations using a computational model that solves the Wason's 2-4-6 task (Wason, 1960). Then we will verify the hypothesis by psychological experiments. Last we will generalize the empirical findings by theoretical task analysis, and discuss why the effect emerges only in the specific situation.

Fundamental issues

Klayman & Ha, in their paper in 1987, gave some decisive answers to several historical questions that had been discussed in the psychological studies using traditional discovery tasks such as the Wason's task (Klayman & Ha, 1987). One of their major conclusions was that there was substantial interaction between the nature of found targets and the effectiveness of hypothesis testing strategies used by subjects. So in this study we will control these two factors in the following experiments.

First, we briefly explain some important concepts about the two factors: the nature of targets that subjects should find and the hypothesis testing strategies that subjects use.

The nature of targets: we categorize targets used in our experiments from the viewpoint of their generality. General targets are defined as the targets, the proportion of whose member (positive instances) to whole instances in the searched space is large. On the other hand, specific targets are defined as the targets, the proportion of whose member is small. For example, the proportion of the instances fitted to "the product is even" and "three evens" to all possible

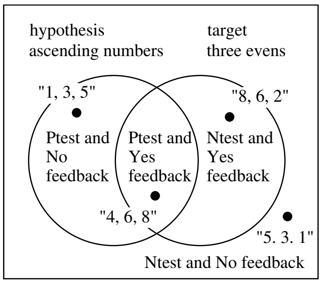
Table 1 Comparisons of the performances in the single, independent, and collaborative conditions in the preceding studies.

	Laughlin (1985)	Laughlin (1986)	Laughlin (1991)					Laughlin (1998)	Freedman (1992)		Okada (1997)
# of group members	4	4	4					4	4		2
task	New Elusis	New Elusis	New Elusis					New Elusis	2-4-6 task		simulated molecular genetics laboratory
single	0.15	0.19	0.06	0.13	0.15	0.14	0.16	0.16	0.33	0.08	1.7
independent	0.47	0.57	0.2	0.3	0.38	0.3	0.41	0.38	0.80	0.28	2.1
collaborative	0.35	0.34	0	0.1	0.2	0.1	0.41	0.34	0.83	0.67	2.9

instances are 7/8 and 1/8 respectively. So the former is an example of a general target and the latter is an example of a specific target.

Hypothesis testing strategies: There are two types of hypothesis testing: a positive test and a negative test. The positive test (Ptest) is conducted by an instance subjects expect to be a target. That is, Ptest is hypothesis testing using a positive instance for the hypothesis; the negative test (Ntest) is hypothesis testing using a negative instance. For example, when a hypothesis is "ascending numbers", hypothesis testing, using an instance, "1, 3, 9", is Ptest; hypothesis testing, using "1, 5, 2", is Ntest.

In the following description, for avoiding the confusion of the basic concepts, we define *Yes* and *No* instances as members and non-members (positive and negative instances) for targets that subjects do not know. On the other hand, we also define *Positive* and *Negative* instances as members and non-members for hypotheses that subjects form. When a subject conducts an experiment using a *Positive* instance for his/her hypothesis, and knows, through the feedback from an experimenter, that the instance is a *Yes* instance for a target, we call that the subject receives *Yes* feedback as a result of his/her *Ptest.*



hypothesis	target: Yes	target: No
Ptest	confirmation	disconfirmation
Ntest	disconfirmation	confirmation

Figure 1 Patterns of confirmation and disconfirmation.

Klayman et. al. summarized the states when a subject's hypothesis was disconfirmed. Figure 1 illustrates those states in the example situation of that the target is "three evens" and the subject's hypothesis is "ascending numbers". When the subject conducts Ptest, using an instance, "1, 3, 5", and receives No feedback, his/her hypothesis is disconfirmed. Another state of disconfirmation is introduced by the combination of Ntest and Yes feedback, using an instance, "8, 6, 2". On the other hand, the states of confirmation are introduced by the combination of Ptest and Yes feedback, using "4, 6, 8", and the combination of Ntest and No feedback, using "5, 3, 1".

Computer simulations

First, we will propose a hypothesis on when advantages of collaboration appear by computer simulations.

Specifications of the model

In the following, we will explain only the summary of our model. Detailed specifications of the model can be seen in our other papers (Miwa, 1999).

The model was constructed on the interactive production system architecture that we had developed for simulating collaborative problem solving processes. The architecture primarily consists of five parts; production sets of System A, production sets of System B, the working memory of System A, the working memory of System B, and a common shared blackboard. 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 has knowledge on the regularities of three numerals. The knowledge is organized as the dimension-value lists. For example, "continuous

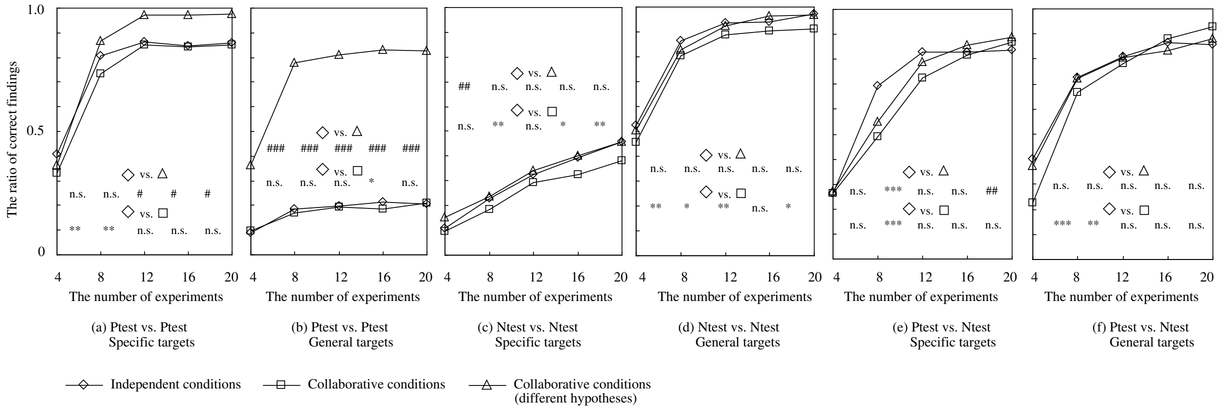


Figure 2 Results of the computer simulations.

evens", "three evens", and "the first numeral is even" are example values of a dimension, "Even-Odd". The dimensions the model uses are: Even-Odd, Order, Interval, Range of digits, Certain digit, Mathematical relationship, Multiples, Divisors, Sum, Product, and Different.

Basically the model searches the hypothesis space randomly in order to generate hypotheses. However, three hypotheses, "three continuous evens", "the 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.

As for a way of hypothesis verification, the principle on when a model's hypothesis is disconfirmed and a next hypothesis is reconstructed is based on the Klayman & Ha's schema shown in the previous section.

The design of simulations

In our computer simulations, we let the two systems find 35 kinds of targets. Examples of the targets are: three continuous evens, ascending numbers, the interval is 2, single digits, the second numeral is 4, first numeral times second numeral minus 2 equals third numeral, multiples of 2, divisors of 24, the sum is a multiple of 12, the product is 48, and three different numbers. The initial instance was "2, 4, 6". For each target, we executed 30 simulations to calculate the percentage of correct solutions.

The computer simulations were conducted based on the following 2 * 3 experimental design.

The nature of targets: We divided the 35 targets into two categories: (a) 17 specific targets and (b) 18 general targets.

Hypothesis testing strategies: Three combinations of hypothesis testing strategies were investigated. They were (a) Ptest and Ptest, (b) Ntest and Ntest, and (c) Ptest and Ntest.

Results of the simulations

Figure 2 shows the results of the computer simulations. The horizontal axis of the figure indicates the number of experiments, that is, the number of generated instances whereas the vertical axis indicates the proportion of correctly finding the 17 specific targets and the 18 general targets.

In Figure 2, the performances in the independent condition and those in the collaborative condition are compared. In the independent condition, we regard that the targets are correctly found when at least one of the two systems, each of which independently tries to find the targets without interaction, reaches the correct solution.

In the collaborative condition, experiments are alternately conducted. Through each simulation, one system generates the half of whole instances; and the other half is generated by the other system. Each experimental result is shared by both two systems, that is, each system knows whole generated instances with Yes or No feedback that is given to each instance.

In addition, the collaborative condition is subdivided into the following two sub-conditions. In one sub-condition, each system simply alternately conducts experiments, not referring to another hypothesis that the other system forms. In this sub-condition, two systems share only the experimental space. In the other sub-condition, one system tries to form a different hypothesis, referring to another hypothesis of the other system. In the latter sub-condition, two systems share the hypothesis space in addition to the experimental space (Klahr &Dunber, 1988).

In the figure, the results of statistical analysis are also indicated. The upper row indicates the difference between the performances in the independent condition and those in the collaborative condition where two systems try to form different hypotheses, whereas the lower row indicates the difference between the performances in the independent condition and those in the collaborative condition where each system does not refer another hypothesis of the other system. The asterisks show the advantage of the independent condition whereas the sharps show the advantage of the collaborative condition. Three levels of significance are used: ### (or ***) for p < .01, ## (or **) for p < .05, and # (or *) for < .1. No significance is indicated with n.s.

Figure 2 indicates that the performance in the collaborative condition exceeds that in the independent condition only when (1) both systems use the Ptest strategy for finding general targets, and (2) both systems try to form different hypotheses, sharing their hypothesis space. In the other cases, the effect of collaboration is not remarkable.

Psychological experiments

To verify the results of the computer simulations in the previous section, we conducted psychological experiments.

Design and procedure

A total of 136 subjects participated in the experiments. Each of them was assigned to each of the following five experimental conditions: (1) the single Ptest condition where a single participant solved a task using Ptest, (2) the single Ntest condition, (3) the collaborative Ptest and Ptest condition where two participants, both of whom were required to use Ptest, collaboratively solved a task, (4) the collaborative Ntest and Ntest condition, and (5) the collaborative Ptest and Ntest condition. Each subject solved two problems. In one problem, "three evens", as a specific target, was discovered. In the other problem, "three different numbers", as a general target, was discovered. The order of the problems was counter-balanced. Twenty-four trials (experiments) were permitted for finding each target. The experimental design is summarized in Table 2.

Table 2 The number of subjects and pairs in each experimental condition.

	single		pair		
	Ptest	Ntest	Ptest v.s. Ptest	Ntest v.s. Ntest	Ptest v.s. Ntest
specific	17(15)	18(14)	16(15)	15(11)	17(11)
general	17(10)	17(12)	17(12)	16(9)	17(9)

In the following discussion, we exclude the results of the subjects who did not follow the experimental instruction requiring to use each hypothesis testing strategy. Table 2 shows the number of subjects (or pairs) assigned to each experimental condition, and, in parenthesis, the number of them who correctly follow the Ptest and Ntest instruction.

Results of the experiments

Figure 3 indicates the experimental results, using the same format of Figure 2. In Figure 3, experimental results that were actually obtained in the experiments are the performances in the collaborative condition (collaborative conditions in (a) through (f)), and those in the single condition where both participants used the same hypothesis testing strategy (single conditions in (a) through (d)). On the other hand, the performances in the single condition, where each subject used a different strategy, are the average scores of the performances in the single Ntest condition and those in the single Ptest condition (single conditions in (e) and

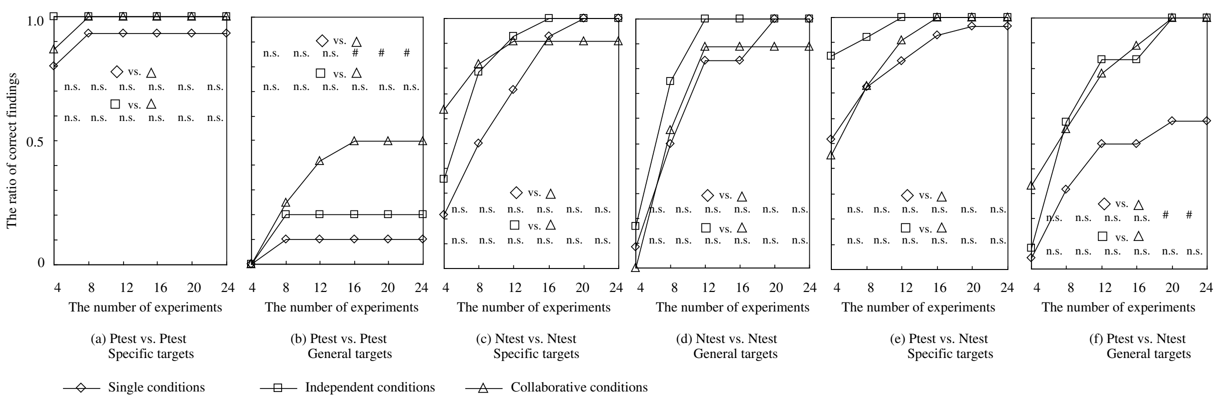


Figure 3 Results of the psychological experiments.

(f)). Additionally the performances in the independent condition are calculated from the performances in the single condition by using the similar procedure indicated in the introduction (independent conditions in (a) and (f)).

As for statistical analysis, the upper row indicates the difference between the performances in the single condition and those in the collaborative condition, whereas the lower row indicates the difference between the performances in the independent condition and those in the collaborative condition.

The statistical analysis shows, in every combination of the hypothesis testing strategies, that the performances in the collaborative condition cannot exceed those in the independent condition. However, only in the combination of Ptest and Ptest for finding the general target, the performance in the collaborative condition exceeds that in the single condition, and a tendency of the advantage of the collaborative condition over the independent condition is observed even though the statistical analysis does not indicate the significant difference.

Next, to confirm the effect of two subjects' forming different hypotheses, we will conduct the following additional analysis. First, we divide the subjects in each collaborative condition into two groups: the subjects who found the correct target earlier and those who did later. The latter group includes those who did not find correct target. Then we measure the average of the proportion of that the subjects in each group

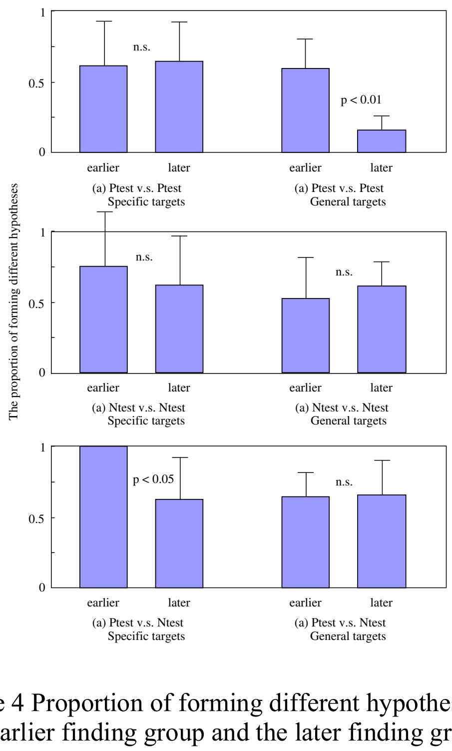


Figure 4 Proportion of forming different hypotheses in the earlier finding group and the later finding group.

maintained different hypotheses through the trials till reaching the solution. Figure 4 shows the result. What we note is that the effect of forming different hypotheses appears in the combination of Ptest and Ptest, especially when finding the general target, whereas this effect does not appear in the combination of Ntest and Ntest. These results are consistent with the findings of the computer simulations.

Theoretical analysis

Why does the advantage of collaboration emerge only when both participants, for finding the general targets, repeatedly conduct Ptest? We will discuss the reason based on the Klayman & Ha's framework of analysis.

Klayman et. al. indicated, by their mathematical analysis, that Ptest was an effective heuristic for finding specific targets; on the other hand, Ntest was effective for finding general targets.

When a target is general, the possibility of receiving Yes feedback is high in the experiments. In the situation, it is difficult that Ptest introduces disconfirmation because the combination of Ptest and Yes feedback introduces confirmation. So Ptest often prevents subjects from finding general targets. In addition, Ntest is an ineffective strategy for finding specific targets because subjects very often receive No feedback as a result of their Ntest. The collaboration of two systems can compensate for these week points of hypothesis testing strategies.

Let us consider the collaborative condition in which both two systems (or two subjects), System A and System B, alternately conduct Ptest, and the systems have different hypotheses. In this situation, it happens that a positive instance for a hypothesis of System A, HA, corresponds to a negative instance for another hypothesis of System B, HB. For example, when a hypothesis HA is "the interval is 2" and a hypothesis HB is "ascending numbers", an instance, "2,

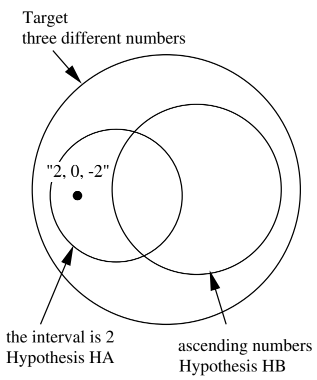


Figure 5 A situation of that Ptest of one system, System A, introduces Ntest of the other system, System B.

0, -2", is this kind of instance (see Figure 5).

When System A conducts Ptest, using this instance, it happens that for System B Ntest is introduced by the instance generated by System A. As a result, Yes feedback introduces disconfirmation of the hypothesis HB because the combination of Ntest for HB and Yes feedback is carried. This brings the effect of collaboration when two systems, both of which use Ptest, find general targets.

An important point is that this function emerges in the interaction between two systems. This advantage is not introduced as the effect of the quantity of the systems. That is, the advantage is not the effect of that the number of systems in the collaborative condition is twice as many as that in the single condition. As you can confirm in Figure 5, when each system independently conducts Ptest, a hypothesis of each system is never disconfirmed. Chances of hypothesis disconfirmation can be introduced only through the collaboration of two systems.

A next question is why this kind of effect does not appear in the combination of Ntest and Ntest when finding specific targets where the probability of subjects' receiving No feedback is very high.

If the above-mentioned type of interaction between two systems emerges in the combination of Ntest and Ntest, the situation in which Ntest of System A introduces Ptest for System B should happen. However, generally speaking, members (positive instances) of a hypothesis is much fewer than the non-members (negative instances). So the possibility of constructing the situation in which Ntest of one system accidentally introduces Ptest for the other system, where the effect of the Ntest and Ntest collaboration appears, is much lower than the possibility of constructing the situation in which Ptest of one system introduces Ntest for the other system, where the effect of the Ptest and Ptest collaboration appears. This is the reason why only the combination of Ptest and Ptest introduces the effect of collaboration.

Conclusions

In the introduction of this paper, we indicated that the effects of collaboration rarely appeared in the psychological experiments, using orthodox simple discovery tasks. We empirically demonstrated a situation in which those effects of collaboration emerged, and theoretically discussed why the effects were introduced. Concretely, we indicated that the effects appeared when both subjects (systems) verified their hypotheses by using Ptest for finding general targets. This result is more interesting, as a finding on collaborative discovery, when we note that humans have a cognitive bias of tending to use Ptest more frequently.

Our empirical findings and theoretical discussions conclude that (1) generally speaking, simply solving a problem together rarely introduces the effects of collaboration, (2) to introduce the effects of collaboration, it is needed that the interaction between collaborative systems brings new abilities, such as a function for introducing disconfirmation of their hypotheses, which are not involved in each individual system, and (3) the possibility of bringing those abilities depends on natures of objects that systems investigate, strategies and heuristics that systems use, and the relation between these factors.

References

Freedman, E. (1992). Scientific Induction: Individual versus Group Processes and Multiple Hypotheses. Proceedings of the 14th annual meeting of cognitive science society, 183-188.

Gorman, M. (1992). Simulating science: heuristics, mental models, and technoscientific thinking. Indiana university press.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211-228.

Laughlin, P.R., & Futoran, G.C. (1985). Collective induction: Social combination and sequential transition. Journal of Personality and Social Psychology, 48, 608-613.

Laughlin, P. R., & McGlynn, R. P. (1986). Collective induction: mutual group and individual influence by exchange of hypotheses and evidence. Journal of Experimental Social Psychology, 22, 567-589.

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: recognition of truth, rejection of error, and collective information processing. Journal of Personality and Social Psychology, 61(1), 50-67.

Laughlin, P. R., Bonner, B. L., & Altermatt, T. W. (1998). Collective versus individual induction with single versus multiple hypotheses. Journal of Personality and Social Psychology, 75 (6), 1481-1489.

Miwa, K. (1999). Collaborative Hypothesis Testing Process by Interacting Production Systems, Lecture Notes of Artificial Intelligence, 1721, 56-67.

Newstead, S., & Evans, J. (Eds.). (1995). Perspectives on Thinking and Reasoning. UK: Lawrence Erlbaum Associates Ltd.

Okada, T., & Simon, H. (1997). Collaborative discovery in a scientific domain. Cognitive Science, 21, 109-146.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly journal of experimental psychology, 12, 129-140.