Collaborative discovery in a scientific domain

Takeshi Okada
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260
takeshi+@pitt.edu

Herbert A. Simon
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
has@cs.cmu.edu

Abstract

This study compares the performance of Pairs with Single subjects in discovering scientific laws with the aid of experiments; describes differences between discovery processes of Pairs and Singles; and identifies the important contributing variables. 36 male undergraduate science majors solved a molecular genetics task in a computer micro-world called the Simulated Molecular Genetics Laboratory (Dunbar, 1993). Pairs were more successful in discovery than Singles and participated more actively in explanatory activities (i.e., entertaining hypotheses, considering alternative ideas and justifications). Explanatory activities were effective only when the subjects also conducted crucial experiments. Explanatory activities were facilitated when paired subjects requested explanations from each other and carefully attended to them.

Scientific discovery through collaboration is a common practice in modern science, but the processes of collaboration have not been well studied. There are several potential approaches to studying collaborative discovery processes, including historical case studies, psychological experiments in the lab, participatory field observation, and interviews with researchers actively involved in collaborative research. Each approach has its own advantages and limitations. As a result, combining several approaches seems most likely to produce a complete and accurate description of the processes of collaborative scientific discovery. In this paper, we will report experimental data based on the Dual Space Search Model of scientific discovery originated by Simon and Lea (1974) and developed by Klahr & Dunbar (1988). Our main focus is on the differences between Singles' and Pairs' discovery processes. We will answer the following questions: (1) Do Pairs perform better than Singles in a scientific discovery task? (2) What are the differences between their discovery processes? (3) What variables are mainly responsible for success?

Method

Design

The experiment has a between-subjects factorial design (Pairs condition vs. Single-subjects condition).

Subjects

Subjects were 27 male science major undergraduates at Carnegie Mellon University. Subjects were randomly assigned to one of the two conditions. Each subject had to bring a friend, who was also a male science undergraduate.

Apparatus and task

A Macintosh computer running the Simulated Molecular Genetics Laboratory (Dunbar, 1993; Dunbar & Schunn, 1990) was used. Using this Laboratory, subjects can learn basic concepts and techniques in molecular genetics, and conduct experiments to discover a scientific mechanism. The detailed explanation of the mechanism to be discovered can be found in Dunbar (1993). A video and two audio tape recorders were used to record the verbal protocols and the computer display.

Procedure

(1) A warm up task was administered to familiarize subjects with giving verbal protocols.

(2) Subjects followed the instructions on the computer display to acquire basic knowledge and techniques in molecular genetics and to do a practice task.

(3) *Discovery phase* Subjects' goal was to find out how I, O, and P genes work to control the beta gene to produce the beta enzyme in a cell. After the first two experiments (conducted as part of the instruction) the experimenter asked the subjects to report and write down their initial hypotheses. Then, in the Pairs condition, subjects were told to reach a consensus on the mechanism by conducting experiments with discussion. In the Singles condition, they were told to find the mechanism by conducting experiments while talking aloud. Both Pairs and Singles were asked to report the discovery, when reached, before the camera, so that other people could review the findings. Each time they conducted an experiment, the computer screen displayed information about their previous experiments. When subjects discovered the mechanism, they reported their conclusions in front of the video camera and then wrote down their conclusions.

Results and Discussion

Quality of performance

To measure outcomes in terms of discovery, we rated performance (i.e., subjects' final hypothesis) on a 5-point-scale. As Table 1 shows, Pairs outperformed Singles ($t(16)=2.69$, $p<.05$), with mean scores of 2.89 and 1.67, respectively.

Table 1 Differences between Pairs and Singles in terms of performance and time measures

Measures	means & (SDs) Pairs		means & (SDs) Singles		p of t tests
Discovery score (full time)	2.89	(0.93)	1.67	(1.00)	<.05
Hypothetical Pairs' (see Singles' column) discovery score	2.89	(0.93)	2.17	(1.08)	=.06
Reported Math SAT scores (Pairs = average score)	706.25	(30.38)	663.33	(97.97)	=.43
Reported SAT scores (Math + Verbal) (Pairs = average score)	1253.00	(23.87)	1246.67	(127.85)	=.92
Solution time (min.)	29.33	(14.85)	23.02	(10.98)	=.32

Table 2 The differences between Pairs and Singles in experimentation measures

	Measures	Pairs' mean and (SD)	Singles' mean and (SD)	p of t-tests
Amount	Number of experiments	13.89 (7.54)	13.89 (6.92)	N.S.
Breadth of E-space search	Dimension search score	11.78 (2.33)	11.44 (2.55)	N.S.
Crucial experiments	% of types of crucial experiments	88.89 (10.54)	86.67 (14.14)	N.S.
VOTAT	Mean feature difference score	1.83 (.14)	1.70 (.29)	N.S.

Reasons for superiority of Pairs over Singles

In this section, we will list plausible hypotheses to explain the superior performance of Pairs over Singles, and test each hypothesis with our data.

Hypothesis (1-1): Members of Pairs happened to have higher intelligence.

To test this possibility, we asked subjects to report their Math and Verbal SAT scores, and found no significant differences between the Singles' scores and the average scores for two members in each Pair (See Table 1).

Hypothesis (1-2): Pairs had better initial hypotheses than Singles.

This hypothesis was not supported by the data. All Singles and all except one Pairs reported an incorrect initial hypothesis.

Hypothesis (1-3): Pairs had twice as great a probability as Singles of getting the right final hypothesis, even without interaction.

We paired all combinations of Singles subjects (9x8/2 = 72 cases). In each case we picked the higher score of the two as the score for a hypothetical pair. Then, we compared the hypothetical pairs' with the real Pairs' mean and SD (See Table 1). Contrary to the hypothesis, the score of the real Pairs (2.89) was better than that of the hypothetical pairs (2.17) (t (43)=1.84, p=.06). Therefore, the superior performance of Pairs seems to depend on the members' interactions instead of just the intelligence, initial ideas, or performance of the abler member.

Hypothesis (1-4): Pairs spent more time than Singles.

Pairs might have been more motivated and spent more time than Singles. However, the difference between them is not statistically significant (t (16)=1.03, p=.32) (See Table 1).

Hypothesis (1-5): Pairs searched the experiment space more effectively.

We investigated subjects' experiment space search processes using various measures adapted from Schunn & Dunbar (in preparation). Table 2 shows the means and SDs for each measure. There were no significant differences between Pairs and Singles.

Hypothesis (1-6): Pairs talked more than Singles.

(See Table 3). Pairs talked, on average, twice as much as Singles throughout the solving process (t (16)=2.73, p<.05). We will discuss below what they talked about.

Hypothesis (1-7): Pairs entertained hypotheses more often than Singles.

Science aims at explaining phenomena. Therefore, how much subjects talked about hypotheses should be a useful measure of scientific discovery. We created two measures: the percent of units[1] in which subjects

[1] We define a unit as the period between two adjacent experiments and also the discussion period following completion of the final experiment.

Table 3 Differences between Pairs and Singles in terms of protocol measures

Measures	means and (SDs) Pairs		means and (SDs) Singles		p of t tests
Number of words	2216.22	(1135.41)	1090.67	(489.90)	<.05
Number of hypotheses	29.56	(13.45)	14.00	(5.10)	<.01
first half	14.56	(5.59)	4.67	(2.60)	<.01
second half	15.00	(9.27)	9.33	(3.43)	=.10
Hypothetical Pairs' number of hypotheses (see Singles)	29.56	(13.45)	28.00	(6.45)	=.61
Number of different types of hypotheses	10.44	(3.24)	7.78	(2.28)	=.06
first half	7.11	(2.76)	2.89	(1.36)	<.01
second half	6.11	(3.55)	6.22	(1.72)	=.93
Hypothetical Pairs' number of types of hypotheses (see Singles)	10.44	(3.24)	13.11	(2.69)	<.05
% of units with summarizing data	48	(21)	47	(13)	=.89
% of units with hypotheses	74	(21)	56	(24)	=.10
% of units with prediction	31	(19)	14	(13)	<.05
% of units with critique	33	(21)	2	(4)	<.01
% of units with alternative hypotheses	25	(26)	6	(10)	=.05
% of units with combined-justification	58	(29)	39	(24)	=.15
% of units with justification through experimental results	41	(19)	24	(11)	<.05
% of units with plan for new experiments to test hypotheses	37	(24)	17	(16)	=.05
% of units with testability of hypotheses	9	(8)	2	(6)	<.05
% of units with justification using several experimental results	35	(19)	19	(13)	=.05
% of units with argument about justification	24	(16)	9	(10)	<.05
Hypothetical Pairs' (see Singles' column) % of units with hypotheses	74	(21)	70	(21)	=.58
Hypothetical Pairs' % of units with alternative hypotheses	25	(26)	11	(11)	<.05
Hypothetical Pairs' % of units with combined-justification	58	(29)	52	(22)	=.49
Hypothetical Pairs' % of units with justification through experimental results	41	(19)	31	(9)	<.05
Hypothetical Pairs' % of units with plan for new experiments	37	(24)	26	(14)	=.08
Hypothetical Pairs' % of units with testability of hypotheses	9	(8)	4	(7)	=.06
Hypothetical Pairs' % of units with justification using several results	35	(19)	26	(14)	=.11
Hypothetical Pairs' % of units with argument about justification	24	(16)	15	(10)	<.05

entertained at least one hypothesis and the absolute number of hypotheses that subjects entertained throughout the entire process. Hypothesis (1-7) was supported by these analyses.

Pairs entertained hypotheses more often than Singles (See Table 3). Although the difference (74% versus 56%) was not significant, it approached the 10 percent level ($t(16)= 1.74$, $p=.10$). Pairs entertained about twice as many hypotheses as Singles (29.56 versus 14.00; $t(16)=3.25$, $p<.01$). The difference was most salient in the first half period (14.56 versus 4.67; $t(16)=4.81$, $p<.001$).

Hypothesis (1-8): Pairs entertained alternative hypotheses more often than Singles.

To convince other members of the scientific community, scientists have to consider alternative plausible hypotheses. Therefore, talking about alternatives may also be important for discovery. (1) We measured the percent of units in which subjects entertained alternative hypotheses. (2) We measured the number of

Table 4 Means (SDs) of performance scores according to occurrence of crucial experiments and explanatory activities

Crucial experiments	Explanatory activities: High	Explanatory activities: Low	Totals
High	3.75 (0.50)	2.00 (1.41)	2.88
Low	1.75 (0.50)	1.83 (0.75)	1.80
Totals	2.75	1.90	

different types of hypotheses that subjects entertained throughout the entire discovery process.

As Table 3 shows, Pairs entertained alternative hypotheses more often than Singles (25 versus 6; t(16)=2.09, p=.05), and also, a greater variety of hypotheses (in total, 10.44 versus 7.78; t(16)=2.02, p=.06). The difference between Singles and Pairs in variety of hypotheses was especially dramatic in the first half of the session (7.11 versus 2.89; t (16)=4.12, p<.001).

> Hypothesis (1-9): Pairs talked about justification more often than Singles.

As scientific claims are accepted or rejected on the basis of evidence, talk about justification may predict success in discovery. One measure of justification is the percentage of units in which subjects talked about justification by experimental results. Another is the percentage of units in which subjects mentioned more than one experimental result to justify their hypotheses. A third measure is the percentage of units in which subjects planned a new experiment to test a hypothesis. A fourth measure is the percentage of units in which subjects talked about the testability of hypotheses. A fifth is the percentage of units in which subjects argued against a justification (partner's or own). As Table 3 shows, all of the measures just described indicate that Pairs considered justification of their hypotheses more often than Singles (1. Justification with results: 41 versus 24; t (16)=1.89, p<.05; 2. Justification with several results: 35 versus 10; t (16)=2.08, p=.05; 3. Experiment to test a hypothesis: 37 versus 17; t (16)=2.07, p=.05; 4. Testability: 9 versus 2; t (16)=2.21, p<.05; 5. Justification with arguments: 24 versus 15; t (16)=2.37, p<.05). We also combined three main measures, 1, 3 and 4, to form a summary measure of justification. Although the difference was not statistically significant, Pairs considered justification more often than Singles (58 units versus 39; t (16)=1.60, p=.13).

Overall, data suggest that an important reason why Pairs performed better is because they participated in explanatory activities such as entertaining hypotheses (hypothesis 2.7), talking about alternative ideas (hypothesis 2.8), and considering justification (hypothesis 2.7) more often than Singles.

Variables that affect performance

To determine whether the differences in behavior we have identified between Pairs and Singles cause the differences in performance, we should inquire how well such measures as entertaining hypotheses and thinking about justification predict performance. Although the number of subjects in each condition was too small for multiple regression analysis with many measures, exploratory regression analyses could suggest which variables were important for discovery. Therefore, we computed simple regressions between performance and each measure in Table 1, Table 2, and Table 3 in each condition separately.

Although the previous analyses would suggest that entertaining hypotheses, considering alternatives, and thinking about justification would predict performance well, there were no significant correlations between performance and the verbal protocol measures (hypotheses, alternatives, and justification). Instead, the experiment space search measures were strong predictors of Pairs' performance, but did not predict Singles' performance at all. The strongest predictor for Pairs' performance was percentage of crucial experiments, which accounted for 78 % of the variance in Pairs' performance (F (1, 7) = 24.11, p<.01), but only 3 % of the variance in Singles' performance (F (1, 7) = 0.23, p=.65).

We propose the following interpretation of these findings:

> Hypothesis (1-10): Due to their active participation in explanatory activities, Pairs could use information from experiment space search effectively in order to make discoveries. On the other hand, Singles could not do so because they did not actively participate in explanatory activities.[2]

In order to check this possibility, we divided subjects according to their scores for percentage of crucial experiments, and their mean scores for explanatory activities (i.e., the combined score of entertaining hypothesis and thinking about justification). Due to the small number of the subjects, Pairs and Singles were combined for this analysis. Table 4 shows that the Pairs and Singles who both conducted all of the crucial experiments and actively participated in explanatory activities outperformed the subjects who did only crucial experiments, only explanatory activities, or neither. An ANOVA shows significant main effects (crucial experiments: F(1, 14)=6.94, p<.05; explanatory activities: F(1,14)=4.11, p=.06) and interaction (F(1,14)=4.97, p<.05). It appears that neither crucial experiments nor

[2] Remember that Pairs' and Single's percentages of crucial experiments were nearly equal (89% versus 87%).

Table 5 Differences between Pairs and Singles in terms of requests for explanation

Measures	means and (SDs) Pairs		means and (SDs) Singles		p of t tests
% of units with requests for explanation	38	(21)	20	(21)	=.10
with requests for description and summary	5	(7)	2	(3)	=.32
with requests for hypothesis	24	(22)	17	(19)	=.46
with requests for justifications	11	(8)	1	(3)	<.01
% of units with answers to requests for explanation (A/S)	80 (n=9)	(19)	44 (n=7)	(37)	<.05
with answers to requests for description and summary (A/S)	100 (n=9)	(0)	100 (n=7)	(0)	-
with answers to requests for hypothesis (A/S)	73 (n=8)	(20)	48 (n=7)	(43)	=.15
with answers to requests for justification (A/S)	78 (n=7)	(40)	65 (n=2)	(21)	-

explanatory activities, by themselves, were enough to lead to discovery.

These data show clearly that entertaining hypotheses and thinking about their justification play important roles in discovery, especially when the experiments are informative. However, before reading a final conclusion about the role of explanatory activities, we should consider whether merely participating in explanatory activities is sufficient or whether collaborative explanatory activities are necessary, for two Singles could produce as much explanatory activity as one Pair.

To test this possibility, we again examined the discussion processes of real Pairs and Hypothetical Pairs. As Table 3 shows, we had found no difference between those two groups in the number of hypotheses generated. In the case of the number of types of hypotheses, the Hypothetical Pairs were even better than the real Pairs. We also checked other discussion measures, using the higher score of two Singles as a Hypothetical Pair's score. As Table 3 shows, in general, the real Pairs entertained alternative hypotheses and participated in justifications more often than the Hypothetical Pairs. As the performance of the real Pairs was better than that of the Hypothetical Pairs, these data suggest that interactive or collaborative explanatory activities are important.

Reasons for differences in explanatory activity

Why did Pairs entertain hypotheses and justifications more often than Singles?

In this section, we will list some plausible hypotheses to account for this difference. Scientific explanations move from a mere description of results, which doesn't generalize beyond the specific case; to a summary of results; to causal explanation of a phenomenon; to justification of the causal explanation. Each level of explanation could be regarded as an answer to a specific question. A description of results could answer: "What was going on?" A summary of results could answer: "How did it happen?" A causal explanation could answer: "Why did it happen?" A justification of explanation could answer: "What evidence supports the explanation?" When people asks themselves these questions, they often utter metacognitive statements such as "I am not sure what it means" and "I wonder how it happened." When we speak of requests for explanation in this paper, we include such metacognitive statements in addition to explicit questions.

Not all explanations are responses to requests for explanation, but such requests may indicate important decision points requiring conscious, reflective thinking. Therefore, we tested the following two hypotheses.

Hypothesis (2-1): Pairs requested explanations more often than Singles.

Table 5 shows that Pairs made more requests for explanation than Singles (38 versus 20; $t(16)=1.77$, $p=.10$). The main difference was found in the requests for justification (11 versus 1; $t(16)=3.12$, $p<.01$); Pairs sometimes questioned whether their hypotheses were justified, while Singles rarely did so.

Hypothesis (2-2): Pairs answered such requests more often than Singles.

Table 5 shows that Pairs answered requests for explanation more often than Singles (80 versus 44; $t(16)=2.59$, $p<.05$). The main difference was found in the answers to requests for hypotheses, although this difference did not reach statistical significance (73 versus 48; $t(13)=1.53$, $p=.15$).

These data suggest that requests for explanation play an important role in producing explanations. Pairs participated in such activities more often than Singles.

General Discussion

Recent studies in cognitive psychology have shown that explanation plays an important role in problem solving and learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1944). Chi and her colleagues pointed out that self-explanation is an important activity when learning from examples and from texts in scientific domains. Good learners, in contrast to poor learners, explained to themselves the meaning of example procedures, or the content of the text, and related the procedures or the content to problem solving goals and higher level content knowledge. The simple intervention of prompting self-explanation improved students' learning dramatically. Although these studies focused on individual learning situations, such explanatory processes seem to be even more important for collaborative learning situations.

Studies of collaborative learning and problem solving in developmental psychology have also found that participation in explanatory activities makes students learn successfully through collaboration (e.g., Brown & Palincsar, 1989; Kruger, 1993; Teasley, 1995). For example, Brown, Palincsar and their colleagues have developed a teaching strategy called "reciprocal teaching" in order to improve students' comprehension skills. The reciprocal teaching strategy, which includes explanatory activities, dramatically improved the students' comprehension as well as monitoring skills. These studies suggest that participating in explanatory activities in social situations also improves students' learning.

All of the recent studies on explanatory activities including the present study converge to tell us that these activities are a crucial component of successful intellectual behavior. Explanatory activities help people to connect pieces of information into an organized theory. Having others as monitors encourages people to participate in such activity, and helps them to construct their theories more actively and more deeply (Miyake, 1986).

Conclusion

Scientific discoveries often are made in social situations, and collaborative research has been emerging rapidly as the predominate form of scientific activity in many domains, but most previous studies in the psychology of science have focused on individual discovery processes. On the other hand, a few recent studies in developmental psychology (see Azmitia & Perlmutter, 1989) and group problem solving (see Hill, 1982; Levine and Resnick, 1993) have examined the processes of collaboration. Our study identifies specific problem solving processes, notably explanatory activities and appropriate data collection, that are important to successful discovery, describes how these processes are accomplished, and shows how they are facilitated by collaboration. It thereby takes an essential first step toward integrating work on the psychology of science with work on the psychology of collaboration in order to capture a broader view of scientific discovery.

Acknowledgment

This research was supported by the A. W. Mellon Foundation under the project title, "The literacy in science research center at CMU." The preparation of this paper was also supported by McDonnell Foundation CSEP Postdoctoral Fellowship for the first author. We would like to thank David Klahr, Jill Larkin, Kevin Crowley, Chris Schunn, and two anonymous reviewers for comments.

References

Azmitia, M. & Perlmutter, M. (1989). Social influences on children's cognition: State of the art and future directions. In Reese, H. W. (1989). *Advances in child development and behavior*. Vol.22. San Diego, CA: Academic Press.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. Resnick (Ed.), *Knowing, learning, and instruction*. New Jersey: Lawrence Erlbaum Associates.

Kruger, A. C. (1993). Peer collaboration: Conflict, cooperation, or both? *Social Development, 2*, 165-181.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding, *Cognitive Science, 18*, 439-477.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17, 397-434.

Dunbar, K., & Schunn, C. D. (1990). The temporal nature of scientific discovery: The role of priming and analogy. In *Proceedings of the Twelfth Annual Meeting of the Cognitive Science Society*. New Jersey: Lawrence Erlbaum Associates.

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91*, 517-539.

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

Levine, J. M. & Resnick, L. B. (1993). Social foundations of cognition. *Annual Review of Psychology*, 44, 585-612.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10*, 151-177.

Schunn, C.D., & Dunbar, K. (in preparation). Priming, analogy, and awareness in complex reasoning. Submitted manuscript.

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

Shrager, J., & Langley, P. (1990). Computational approaches to scientific discovery. In J. Shrager & P. Langley (Eds.). *Computational models of scientific discovery and theory formation*. San Mateo, California: Morgan Kaufmann.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*, 207-220.