Japanese Psychological Research
1996, Volume 38, No. 2, 74–84

A new experimental method to identify the process of logical reasoning

MASASI HATTORI[1] and MASANORI NAKAGAWA[2]
Department of Behavioral Science, Faculty of Letters, Hokkaido University, Kita-10, Nishi-7, Kita-ku, Sapporo 060, Japan

Abstract: A new method to identify the process of logical reasoning is presented. In spite of its indispensability and importance, we have had few methods to identify a subject's reasoning process, except that of using verbal protocol data. In this paper, for the purpose of objective identification of the reasoning process, we propose a new method to obtain the subject's reasoning process, in terms of a resolution tree for a task of which the logical structure can be written by first-order predicate logic. The results of an experiment using this method are presented. They revealed some interesting features of human reasoning such as, large differences between subjects, remarkable parallel processes, and the existence of subgoals for each subject.

Key words: reasoning, first-order predicate logic, problem solving, resolution tree.

In this paper, a method to identify the process of human logical reasoning is proposed. Suppose a subject is given a logical problem, such as the following (modified from Chang & Lee, 1973, p. 89), and is asked to judge whether the conclusion is true or not.

> *Premises*: The customs officials searched everyone who entered this country who was not a VIP. Some of the drug pushers entered this country and they were searched only by drug pushers. No drug pusher was a VIP.
> *Conclusion*: Some of the officials were drug pushers.

Generally, the procedure of derivation of proof from a given set of logical formulae is nondeterministic, and automated procedures often give a vast number of redundant inference steps, even if the problem appears quite simple to human beings. Several efficient methods have been proposed in automated theorem-proving systems of artificial intelligence (AI). There emerges a question: "What is the real inference process of human beings?" In order to answer this question, it is important to know the process of human logical inference for problems like that given above, in other words, to find out how a human being solves logical problems in what steps of inference, and to know where (s)he infers correctly and where (s)he infers incorrectly, and so on, throughout the whole logical reasoning process.

Though many models of deductive reasoning have been developed in psychology (for reviews, see, e.g., Evans, Newstead, & Byrne, 1993; Johnson-Laird & Byrne, 1991), we have had few methods to identify the subject's reasoning process except that of using verbal (oral or written) protocol data – the method of *protocol*

[1] The authors would like to thank Mark Radford of the International Survey Research Corporation, and two anonymous reviewers for their critical reading of manuscripts and valuable comments on an earlier draft. This paper was presented in part at the 8th Annual Meeting of the Japanese Cognitive Science Society, Tokyo, July 1991.
Correspondence and requests for reprints should be sent to Masasi Hattori.
[2] Now at Department of Education, Faculty of Engineering, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152, Japan.

analysis (e.g., Ericsson & Simon, 1984). Even though this method is useful in many cases, there are some problems with it in strictly and objectively identifying reasoning processes. For example, the subjects must be specially trained in order to obtain useful protocol data. Second, subjects may not always tell the truth. Third, subjects cannot always verbalize all the processes. Fourth, it leaves room for involving the experimenter's subjectiveness in interpreting the data, and so on.

We propose, in this paper, a method to identify objectively the process of human logical reasoning. This method has several restrictions at present, such as the logical structure of the task being represented by first-order predicate logic, as is detailed in the next section. On the other hand, the method of protocol analysis may be a generic one for a variety of cognitive tasks. Therefore, if the present method is used simultaneously with protocol analysis, it may lead to a new paradigm for experimental studies of logical reasoning. The discussion about the extension of this method and general problems in methodology will be presented in the last section.

Representation of the problems and solutions

Below, we propose a method to identify a subject's reasoning process when (s)he is solving logical problems. We introduce the notion of logical formulae and graphs in order to represent given problems and subjects' solutions of these problems.

The method we propose here is based on the resolution principle (Robinson, 1965), which is one of the algorithms of automated theorem proving that verify the validity of a given formula of first-order predicate logic. For instance, the logical structure of the problem given above is represented by the following *well formed formulas* (wffs). Let $A(x)$ mean "x entered this country," $B(x)$ mean "x was a VIP," $C(x)$ mean "x is a human," $D(x,y)$ mean "y searched x," $E(x)$ mean "x was a customs official," and $F(x)$ mean "x was a drug pusher." Then the premises can be represented by the conjunction of the following wffs:

F_1: $\forall x[[A(x) \wedge \neg B(x) \wedge C(x)] \rightarrow \exists y[D(x,y) \wedge E(y)]]$,
F_2: $\exists x[F(x) \wedge A(x) \wedge \forall y[D(x,y) \rightarrow F(y)]]$,
F_3: $\forall x[F(x) \rightarrow \neg B(x)]$,
F_4: $\forall x[F(x) \rightarrow C(x)]$,

and the conclusion is

$[G: \exists x[F(x) \wedge E(x)]]$.

Generally, as a matter of convenience, automated theorem-proving procedures including the resolution are not based on validation but on refutation. These procedures are applied to a standard form of a wff. The conjunction of the premises and the negation of the conclusion should be transformed into *prenex normal form* whose matrix is in *conjunctive normal form*, $Q_1x_1Q_2x_2 \cdots Q_nx_n\ [F_1 \wedge \cdots \wedge F_m]$, where every Q_ix_i, $i = 1, \ldots, n$, is either $\forall x_i$ or $\exists x_i$, and every F_j, $j = 1, \ldots, m$ is a disjunction of literals – a literal means an atom or the negation of an atom (e.g., $A(a)$, $\neg B(x)$, etc.). Furthermore, without affecting the inconsistency property, the existential quantifiers in the prefix of this formula can be eliminated. Suppose Q_r is an existential quantifier in prefix $Q_1x_1Q_2x_2 \cdots Q_nx_n$, $1 \leqslant r \leqslant n$. If no universal quantifier appears before Q_r, we choose a new constant, c, and delete Q_rx_r from the prefix. If $Q_{s1}, Q_{s2}, \cdots, Q_{sm}$ are all the universal quantifiers appearing before Q_r, $1 \leqslant s_1 < s_2 < \ldots < s_m < r$, we choose a new m-place function symbol f, replace all x_r by $f(x_{s1}, x_{s2}, \ldots x_{sm})$, and delete Q_rx_r from the prefix. The function introduced here is called *Skolem* function. The wff transformed into standard form through the above process becomes a conjunction of *clauses* governed only by the universal quantifier, where a clause is a finite disjunction of zero or more literals. Thus, for convenience, omitting the universal quantifier and conjunction symbol, we represent a wff by a set of clauses like below.

$$S = \{\neg A(x) \vee B(x) \vee \neg C(x) \vee D(x, f(x)), \quad (1)$$
$$\neg A(y) \vee B(y) \vee \neg C(y) \vee E(f(y)), \quad (2)$$

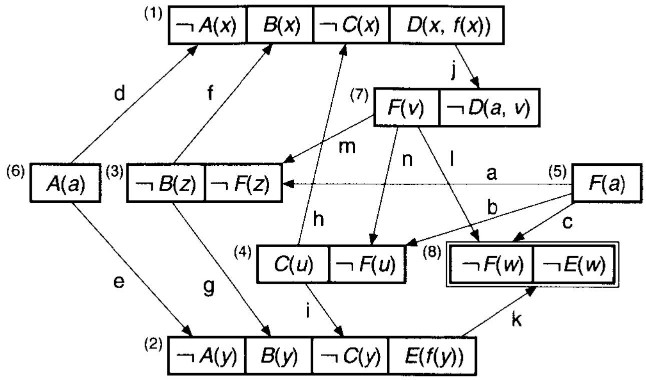


Figure 1. An example of a connection graph. The node partition corresponding to the conclusion (i.e., top clause) is written in the duplicate line box.

$$\begin{array}{lr} \neg F(z) \vee \neg B(z), & (3) \\ \neg F(u) \vee C(u), & (4) \\ F(a), & (5) \\ A(a), & (6) \\ \neg D(a, v) \vee F(v), & (7) \\ \neg F(w) \vee \neg E(w)\}. & (8) \end{array}$$

Here, (1) and (2) are obtained from F_1, (3) is obtained from F_3, (4) is obtained from F_4, (5), (6), and (7) are obtained from F_2, and (8) is obtained from negation of G.

In order to clarify the logical structure of the task, we represent the set S of clauses by a (directed) connection graph[3], as shown in Figure 1 (Chang & Slagle, 1979; Sickel, 1976, 1977). Every edge in the connection graph connects the literals that are potentially complementary with each other. It corresponds to the possible refutation and substitution in the set of clauses.

The central idea of the resolution procedure is elimination of literals in the set of clauses: for any two clauses C_1 and C_2, if there is a literal L_1 in C_1 that is complementary to a literal L_2 in C_2, delete L_1 and L_2 from C_1 and C_2 respectively, and construct the disjunction of the remaining clauses. The constructed clause is called a *resolvent* of C_1 and C_2 (see Chang & Lee, 1973, Chapter 5). A step in the resolution in the set of clauses corresponds to an edge in the connection graph. Figure 2 shows an example solution of the problem in terms of a resolution tree.[4] Every branch pair in this figure represents a step in the resolution corresponding to the edge in Figure 1, and the label of each pair of branches and that of the edge are named identically. To take a simple example of resolution (Figure 2), the step in the resolution labeled "a" deletes literals $F(a)$ and $\neg F(a)$ from two clauses, (5) and (3), and the resolvent is $\neg B(a)$, labeled α. We will show in the following section that a subject's reasoning process can correspond to a resolution tree.

Practical design of the experiment

To identify a reasoning process with the resolution tree, we designed an experiment as follows.

Framework

The experimental task is a quasichemical experiment conducted on a color video display controlled by a personal computer. The task need not necessarily be a chemical experiment; as long as the task satisfies the methodological

[3] The original version of this graph is called *clause interconnectivity graph* in Sickel (1976, 1977).

[4] This solution is based on the method of SPU (selective positive unit resolution; see Nagao & Fuchi, 1983, p. 105).

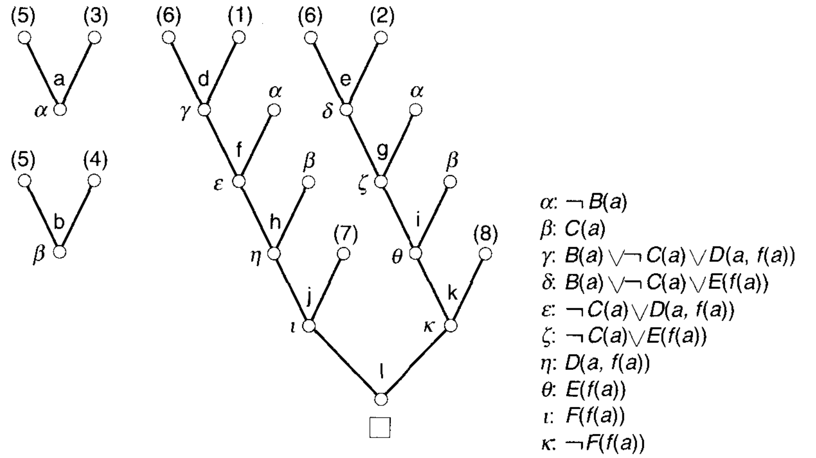


Figure 2. A resolution tree for the example in Figure 1. Each number in this figure indicates a clause in the set S of clauses, and corresponds to that in Figure 1. The label marked on each pair of the branches corresponds to that of the edge in Figure 1.

conditions described below, the style of task is free.

Chemical agents are displayed on the screen, and subjects are given a set of *tests* with which to examine the properties of the agents. Every subject is also given a sheet of paper on which are written a set of *rules* regulating the relations between the properties of the chemical agents. Each rule is written in natural language, and expressed by conditionals. The goal of the problem is to find out which agent satisfies the designated chemical property. Subjects are instructed to solve the problem logically by examining the properties of the chemical agents through the tests and by making inferences from the results of the tests and the given rules. However, in order to execute a test, the following conditions have to be satisfied. These conditions will be called the *bottom-up restriction*.

1. The property of the chemical agent that is to be examined by the test must not be included in any of the consequents of the given rules. (Note that unit clause, e.g., $P(a)$, or $Q(x)$, is not involved.)
2. If the property violates the above condition (i.e., if the property of the agent to be examined by the test is included in the consequent(s) of one or more rules), then all the antecedents of the rule or rules must have been confirmed by preceding tests.

With this restriction, you can represent the structure of the logical task as a connection graph and represent the real reasoning process of each subject as a resolution tree.

Reasoning process and execution of tests

In this experiment, each rule corresponds to a clause in the set S of clauses (it is called *node partition* on a connection graph), and each test corresponds to an edge that connects the complement literals to each other in the connection graph. (In fact, as it will be shown below, a test may correspond to more than one edge.) From the standpoint of formal logic, a chemical agent corresponds to a nonlogical constant. And the results of an experiment will reveal the subject's logical reasoning process objectively as

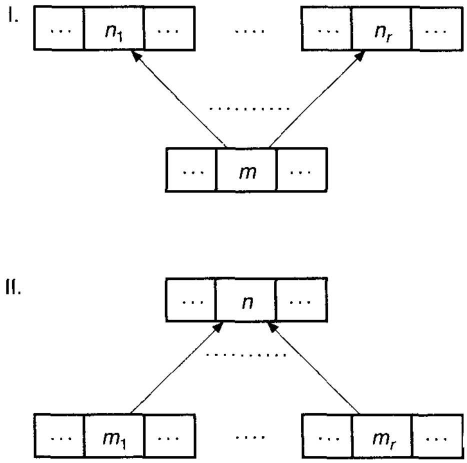


Figure 3. Two cases in which a node has multiple edges. The difference between cases I and II is the direction of the edges. In these cases, the correspondence between test and edge can not be one to one.

a sequence of executed tests, corresponding to the sequence of labels of the edges. In fact, the sequence of executed tests indicates a proof by the refutation procedure on the basis of the resolution principle.

Now, consider some problematical forms of connection graphs. If a node has multiple edges (Figure 3), the relation between a test and a step in the refutation cannot have a one-to-one correspondence. For this reason, there is a need for an additional procedure. In the new procedure, the subject has to answer the following two or three questions, every time (s)he executes the test.

1. "Which rule's condition[5] do you want to satisfy by the test that you intend to conduct at present?"
2. "If you are confident about the result of the test by the given rule, then answer whether the chemical agent *has* or *does not have* the property; if unsure, then answer *uncertain.*" (The answer should be "yes," "no," or "uncertain.")
3. If the subject answered "yes" or "no" to the previous question, (s)he was given the following additional question: "From which rule did you predict the property of the agents?"

The subject is allowed to select more than one rule in question 1, but only one in question 3.

The answer to question 1 resolves the uncertainty of the correspondence between a test and a step in the refutation shown in Figure 3-I, and the answer to question 2 resolves the uncertainty shown in Figure 3-II. Moreover, the answer to question 2 indicates whether the execution of the test is merely a means to find out the agent's property, or whether it corresponds to a step in the refutation. For example, conducting test A under the condition given by the rule "There exists chemical agents that have property A" may not correspond to a step in the refutation, but rather be a means of examining the property. On the other hand, when the rule is "If a chemical agent has property B, it has property A" and it is already evident from the results of test B that the agent has property B, then the execution corresponds to a step in the refutation. This will be substantiated by the answers obtained from the questions.

Experiment

Following the method above, we conducted an experiment to identify the reasoning process of subjects during the solving of a logical task.

Method

Task. Seven rules and a supplementary rule were prepared. The rules and the goal were printed on a sheet of paper and handed to the subjects. They were as follows[6] (all properties in the rules are fictional):

1. *Iso-group* chemical agents which are *polymodific* and not *breakabilic* are *heatable.*

[5] As a matter of fact, the strict meaning of the word "condition" used here is "antecedent."

[6] They were written in Japanese in the experiment: *iso-group* was described with Japanese artificial word *iso-zoku, breakabilic* with *kai-sei, polymodific* with *juuka-sei, drastine* with *geki-sei, paratic* with *para-sei.*

2. *Iso-group* chemical agents which are *polymodific* and not *breakabilic* are *drastine* after *heating*.
3. *Paratic* chemical agents are not *breakabilic*.
4. *Paratic* chemical agents are *polymodific*.
5. There exists a *paratic* chemical agent.
6. There exists an *iso-group* chemical agent.
7. There exists a *paratic* chemical agent after *heating*.

Some chemical agents satisfy rules (5), (6), and (7) simultaneously.

And the goal was:

8. If there are chemical agents that are *paratic* and *drastine*, please find one; if you cannot find any, then form it (by heating) from the given chemical agents.

Let $A(x)$ denote "x is *iso-group*," $B(x)$ denote "x is *breakabilic*," $C(x)$ denote "x is *polymodific*," $D(x,y)$ denote "x changes to y by *heating*," $E(x)$ denote "x is *drastine*," $F(x)$ denote "x is *paratic*." The logical structure of these rules is identical to the example mentioned at the beginning, and the premises and the negation of the conclusion are shown by the set of clauses (1), (2), ..., (8) given under "Representation of the problems and solutions," while the connection graph of these clauses is shown in Figure 1.

The set of clauses pertaining the task contains one Skolem function relating to the predicate S; the Skolem function corresponds to the operation *heating*, and the predicate S denotes the relationship between the agent to be heated and the one generated from it by the operation.

Subjects were given eight chemical agents fictionally named on the video display. In order to avoid a possible selection-order difference among subjects, the properties of chemical agents were not defined in advance but were defined by the subject's selection order itself. For instance, both for a subject who selected the left-most agent first and for another subject who first selected the agent third from the left, the property of the chemical agent was the same. Hence, in whatever order they selected the agents, all the subjects had to conduct the same minimum number of tests that was decided upon before the experiment by the experimenter. Since the goal chemical agent formed by heating is necessarily selected seventh in this experiment, everyone had to test at least six agents before the goal was reached, regardless of selection order. Figure 4 shows an example of the video display from the experiment.

Subjects. Eighteen subjects participated in this study. All subjects were university students, and were native Japanese speakers. Although we did not restrict subjects by their majors, no one was a mathematics major nor a logic major.

Results and discussion

The resolution trees obtained from subjects (two of them are shown in Figures 5 and 7) indicate large variation, showing that differences in people's logical reasoning styles are quite large.

Figure 5 shows a resolution tree for a subject which represents a typical process of reasoning in this task. The tree shows that the first step of the subject's inference is to obtain both clause $B(a) \vee \neg C(a) \vee E(f(a))$ (labeled α in Figure 5, and we call it resolvent α) from rules (2) and (6), and clause $B(a) \vee \neg C(a) \vee D(a, f(a))$ (resolvent β) from rules (1) and (6) concurrently. Repeating the resolution like this, the subject was able to prove (i.e., solve) the problem in seven steps at the last stage of the task.

Some of the steps in the resolution (i.e., the first, second, fourth, and sixth steps) in Figure 5 can be regarded as a kind of parallel resolution. For example, the resolution of the first step is regarded as deriving two resolvents (i.e., α and β) concurrently from three clauses (i.e., rules (1), (2), and (6)), while the ordinary resolution procedure practiced in artificial intelligence derives only one resolvent at a time. We call this parallelism *divergent parallelism*, and it can be schematized generally as Figure 6-I.

The salient feature of this tree is the portion for deriving the resolvent ζ and θ (the fourth and the sixth step, respectively) in the figure. Three or more edges come together at these nodes. This derivation can be regarded as a kind of parallelism. We call it *convergent parallelism*,

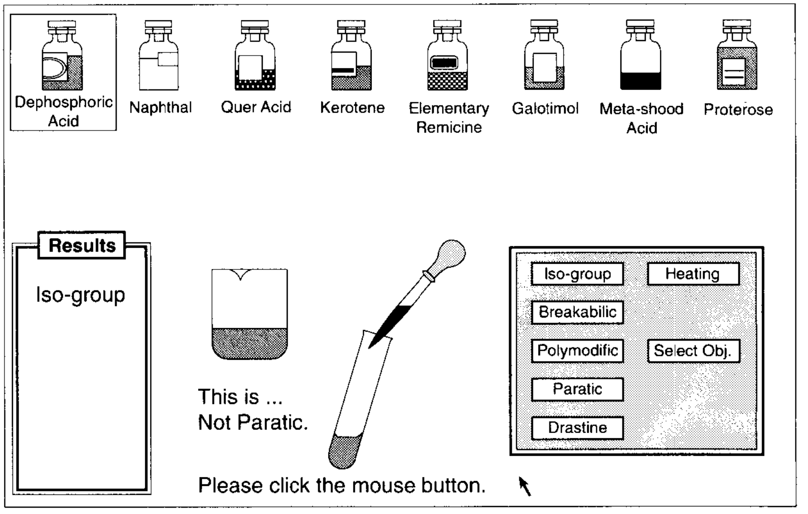
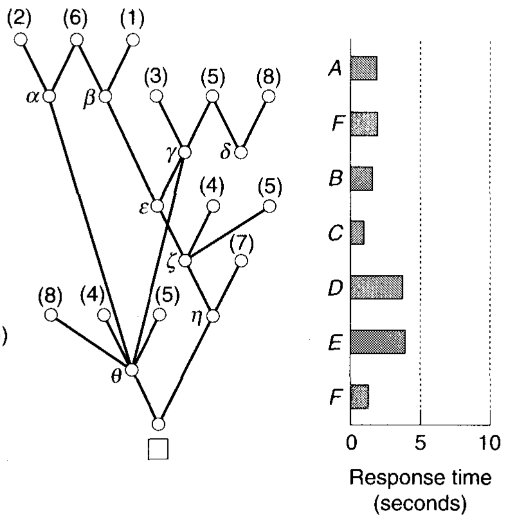


Figure 4. A sample screen from the experiment. This figure indicates an execution of *paratic* test for the extreme left chemical agent, and the result shows that it is not *paratic*. All of the subjects' operations (e.g., selecting a chemical agent from eight, executing a test, etc.) were conducted by clicking a mouse button. The center part of the screen demonstrates each test execution, and varies with the test. The lower left window lists all the results of tests previously executed for the current agent. In the experiment, all messages and all names of the chemical agents and properties on the screen were in Japanese.

and Figure 6-II shows the general form of this parallelism in the resolution tree. These parallelisms are different from "or parallelism" or "and parallelism," which are major concepts of logic programming.

In relation to the response time in Figure 6, while it seems that there is no significant difference between steps, the second or the third step from last takes a little longer. The second step from last corresponds to the execution of test E for the chemical agent after heating. Table 1 shows the mean response time for all subjects. This table reveals that test E took a long response time. The initial response time – time from the beginning of the task to the initial response – is not shown in Table 1 because it was often very long and was regarded as being somewhat different from the other responses. The point to notice is that test E was executed right after test D (heating operation) by the subject of Figure 5 and many others. Test E must be executed after heating because the literal $E(f(y))$ of rule (2) contains a Skolem function. (See formula (2) above.) For this reason, there is a possibility that heating was a kind of subgoal for most subjects, and when they reached the subgoal, they needed extra time for new planning to attain the final goal.

Some subjects derived redundant resolvents. Figure 7 indicates a resolution tree which contains a redundant subtree (represented by broken lines). In Figure 7, we cannot overlook the fact that there was a marked increase in response time on tests B and E (i.e., the seventh and the ninth steps, respectively). With regard to the resolvent κ, the subject concurrently derived resolvent λ, which is indispensable, and μ and ν, which are redundant. The response at the seventh step was right after this derivation, and the ninth step occurred when the subject ended the derivation of redundants. A possible explanation for this round-about means of attaining the goal might be as follows: the subject thought about something for a while, made a false plan and derived a redundant resolvent, ξ (the seventh step), and successively



α: $B(a) \vee \neg C(a) \vee E(f(a))$
β: $B(a) \vee \neg C(a) \vee D(a, f(a))$
γ: $\neg B(a)$
δ: $\neg E(a)$
ε: $\neg C(a) \vee D(a, f(a))$
ζ: $D(a, f(a))$
η: $F(f(a))$
θ: $\neg F(f(a))$

Figure 5. A resolution tree for the task. Each number labeled on the node in this figure (i.e., (1), (2), ..., (8)) denotes a rule number, and each label (i.e., α, β, ..., θ) denotes a resolvent that the subject derived. Note that this tree represents only the last stage of the subject's reasoning process for the task. At this last stage, the subject dealt with the chemical agent that satisfied the conclusion, namely, the last agent that (s)he used in the task. The subject had carried out similar inferences before this stage. The right-hand horizontal bar chart shows response time corresponding to the steps of the inferences. On the vertical axis, the predicate symbol of the test that corresponds to the step of resolution of the left-hand tree is shown. The salient feature of this tree is the portion for deriving the resolvent ζ and θ, which is called convergent parallelism.

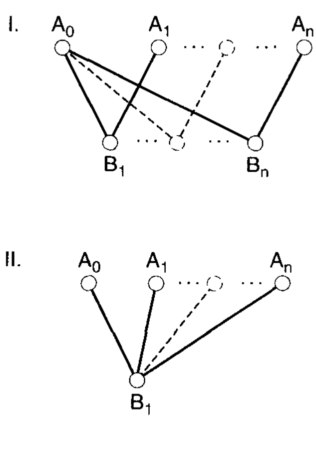


Figure 6. I. Divergent parallelism in resolution tree ($n \geqslant 2$). II. Convergent parallelism in resolution tree ($n \geqslant 2$).

derived o (the eighth step) following this plan. Then the subject became aware of the mistake, thought about it for a while, and then attempted to resolve the contradiction (the ninth step).

General discussion

The experiment reported in this paper revealed details of each subject's reasoning process. The inference process was identified objectively in the form of a partial or whole resolution tree, and response time in each step was informative of the characteristics of the reasoning process. Thus the differences between individuals in reasoning were concretely and objectively identified, and, in particular, the existence of a parallelism in logical inferences was shown. Though further experimental research

Table 1. Mean response time and the total number of executions of the tests for all subjects

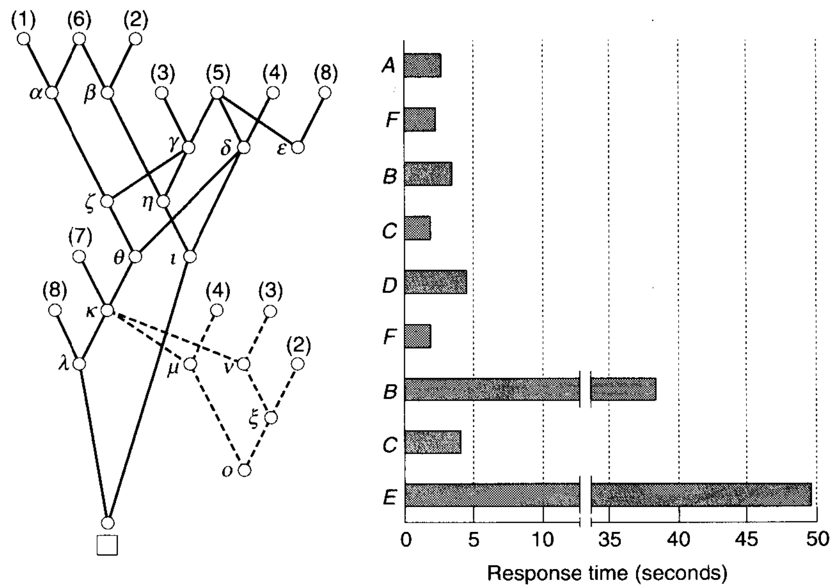
Test	A	B	C	D	E	F
Response time (s)	14.5	11.3	10.5	12.0	31.1	15.5
Numbers of executions	66	49	48	39	23	120

The initial response time was excluded because the time was often very long and appeared somewhat different from the other response times. It is evident that test E took a comparatively long response time.

is needed to confirm the parallelism, it may lead to a new conceptualization of the reasoning process.

Some subjects made redundant steps: part of the reasoning steps of some subjects turned out to be a redundant subtree which had no bearing on the attainment of the goal. But this redundancy differed from that usually encountered in artificial systems. The redundant resolvents derived by automated theorem-proving systems are extensive, diverse, and out of control, while redundant steps of human beings are characterized by their response time. Thus, the comparison of redundancy in human and mechanical systems may provide new insights both for artificial intelligence and for the psychological study of reasoning.

The method we presented has potential for generalization. First, the nature of tasks is not constrained – the method is independent of the superficial meaning of the task, as described above. Second, we can examine all varieties of



α: $B(a) \vee \neg C(a) \vee D(a, f(a))$
β: $B(a) \vee \neg C(a) \vee E(f(a))$
γ: $\neg B(a)$
δ: $C(a)$
ε: $\neg E(a)$
ζ: $\neg C(a) \vee D(a, f(a))$
η: $\neg C(a) \vee E(f(a))$
θ: $D(a, f(a))$
ι: $E(f(a))$
κ: $F(f(a))$
λ: $\neg E(f(a))$
μ: $C(f(a))$
ν: $\neg B(f(a))$
ξ: $\neg A(f(a)) \vee \neg C(f(a)) \vee E(f(f(a)))$
o: $\neg A(f(a)) \vee E(f(f(a)))$

Figure 7. A resolution tree that contains a redundant subtree. The subtree represented by broken lines (which involves resolvent μ, ν, ξ, and o) is unnecessary to finish the task. There is marked increase in response time on tests B and E.

logical structure (at least in terms of classical logic) that correspond to an unsatisfiable set of clauses. Therefore, an experimenter can include any logical structure that (s)he wants to examine in the cognitive task.

On the other hand, there are several problems that need to be resolved. The foremost is the bottom-up restriction (see above). Since many strategies that are not bottom-up, for example, simple SNL (selective negative linear resolution; see Nagao & Fuchi, 1983) can also be considered as a solution for a set of clauses, the subjects should be allowed to select such strategies. The bottom-up restriction of the present experiment may have overly constrained the inference style of subjects.

A closely related problem is that if a subject used a procedure that is not bottom-up, the response was regarded as a mistake and was inhibited thereafter, so that the inference steps that may have occurred after the response failed to be recorded.

Third, though we interpret a rule in the set of clauses as a conditional, for example, $\neg A \vee B$ as "If A then B," this is not a unique interpretation. In classical logic, the formulas $\neg A \vee B$, $A \rightarrow B$, and $\neg B \rightarrow \neg A$ are equivalent to each other. Thus the rule could be expressed by a sentence like "Not A or B (or both)," or "If not B then not A," and so on. The crucial point is whether these sentences are cognitively equivalent or not. If they are not equivalent, it is important to know their effects on reasoning.

Lastly, and with some relevance to the last problem, there is a question concerning whether the set of clauses are sufficient as tools to represent the human reasoning process. It is true that any wff can be transformed into a set of clauses without affecting the inconsistency property, but there remain some problems in the transformation. Let us return to the example at the beginning of this paper. Consider the following two sentences.

1. Some drug pushers entered this country and *all* of them were searched only by drug pushers.
2. Some drug pushers entered this country and *some* of them were searched only by drug pushers.

These two sentences are expressed by following wffs, respectively:

$$\exists x[F(x) \wedge A(x)] \wedge \forall y[F(y) \wedge A(y) \rightarrow \forall z\,[D(y,z) \rightarrow F(z)]]$$

$$\exists x[F(x) \wedge A(x)] \wedge \exists y[F(y) \wedge A(y) \wedge \forall z\,[D(y,z) \rightarrow F(z)]]$$

These wffs, however, are transformed into the same set of clauses:

$$\{F(a), A(a), \neg D(a,z), F(z)\}.$$

When these wffs are transformed into one and the same set of clauses, some information (but not the inconsistency property) is lost. The distinction of meaning between the above two sentences disappears as soon as they are transformed into a set of clauses. It is certain that the above two sentences are different in meaning for human beings. If a difference like this is really important for the human reasoning process, the sets of clauses are insufficient as tools to represent the logical tasks and reasoning process, and need improvement. They also need extension in order to examine other kinds of reasoning processes, which have the characteristics of inductivity, abductivity, nonmonotonicity, and so on. And with respect to the experimental findings reported here, we need to find out in further research whether they are dependent on the personality characteristics of the subjects and/or the nature of the method employed.

References

Chang, C.-L., & Lee, R. C.-T. (1973). *Symbolic logic and mechanical theorem proving*. London: Academic Press.

Chang, C.-L., & Slagle, J. R. (1979). Using rewriting rules for connection graphs to prove theorems. *Artificial Intelligence*, **12**, 159–178.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Evans, J. St B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nagao, M., & Fuchi, K. (1983). *Ronri to imi [Logic and meaning]*. Tokyo: Iwanami.

Robinson, J. A. (1965). A machine-oriented logic based on the resolution principle. *Journal of the ACM*, **12**, 23–41.

Sickel, S. (1976). A search technique for clause interconnectivity graphs. *IEEE Transactions on Computers*, **C-25**, 823–834.

Sickel, S. (1977). Formal grammars as models of logic derivations. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77), Vol. 1* (pp. 544–551). MIT, International Joint Conferences on Artificial Intelligence.

(Received June 27, 1994; accepted March 5, 1995)