Comprehension and Analysis of Information in Text:
II. Decision Making with Texts

Ely Kozminsky, Walter Kintsch, Paul Coren and Lyle E. Bourne, Jr.
Department of Psychology
University of Colorado

Technical Report No. 89-ONR
Institute for the Study of Intellectual Behavior
University of Colorado
Boulder, Colorado 80309
December 1979

This research was sponsored by the Personnel and Training Research Programs, Psychological Science Division, Office of Naval Research, under contract No. N00014-78-C-0433, Contract Authority Identification Number NR 157-422

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.
# Comprehension and analysis of information in text: II. Decision making with texts

**Authors:**
Ely Kozminsky, Walter Kintsch, Paul Coren, Lyle E. Bourne, Jr.

**Performing Organization Name and Address:**
ISIB
University of Colorado
Boulder, Colorado 80309

**Controlling Office Name and Address:**
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

**Report Date:**
December, 1979

**Number of Pages:**
57

**Distribution Statement (of this Report):**
Approved for public release; distribution unlimited

**Security Classification (of this Report):**
Unclassified

**Abstract:**
Decision making based on information in texts was studied in a laboratory analogue of a complex, natural, information-analytic domain. Subjects acting as stock brokers acquired a conjunctive decision rule for predicting the market performance of a fictitious stock. Subjects read quarterly reports containing information on six market-information categories, of which only two were relevant to correct decisions. Decision performance differentiated between Learners (subjects who discovered the relevant...
categories) and Nonlearners. Hypothesis selection behavior was similar to that reported with simpler concept learning problems. The category recall pattern reflected hypothesis selection, decision behavior, and subjects' overall category identification strategies. Further, these data were congruent with a model of text comprehension. Effective decision making in this task was viewed as the ability to acquire an appropriate control schema to guide comprehension and analysis of complex, often unreliable text inputs.
Comprehension and Analysis of Information in Text:

II. Decision Making with Texts

Ely Kozminsky, Walter Kintsch, Paul Coren and Lyle E. Bourne, Jr.

University of Colorado
Abstract

Decision making based on information in texts was studied in a laboratory analogue of a complex, natural, information-analytic domain. Subjects acting as stock brokers acquired a conjunctive decision rule for predicting the market performance of a fictitious stock. Subjects read quarterly reports containing information on six market-information categories, of which only two were relevant to correct decisions. Decision performance differentiated between Learners (subjects who discovered the relevant categories) and Nonlearners. Hypothesis selection behavior was similar to that reported with simpler concept learning problems. The category recall pattern reflected hypothesis selection, decision behavior, and subjects' overall category identification strategies. Further, these data were congruent with a model of text comprehension. Effective decision making in this task was viewed as the ability to acquire an appropriate control schema to guide comprehension and analysis of complex, often unreliable text inputs.
In semantically rich natural environments, human beings are continuously confronted with the necessity to analyze a complex informational input, to identify those elements pertinent to some predefined goal, and to decide on a basis on actions to attain their goals. A variety of familiar tasks exemplify this general process, ranging in complexity from political policy analysis to deciding which TV programs to watch. More often than not, information relevant to the decision task appears at least in part in text form. Further these texts characteristically are low in signal-to-noise (relevant-to-irrelevant information) ratio, riddled with redundancies, inexplicit, and something less than perfectly reliable. In many task domains, such as scientific research or intelligence analysis, the input is massive as well as imprecise. Yet for any goal-directed progress to be achieved, decisions must be made and executed on the basis of critical information derived from the input. It is the purpose of this paper to address, both empirically and theoretically, some of the issues inherent in this complex information-analytic process.

Our specific goal in the study to be reported was to develop and evaluate a simple theoretical model for analytic information processing in a laboratory analogue of one familiar natural domain, viz., stock market analysis. The model rests heavily on the concept of a schema. We view a schema as a
cognitive structure consisting of declarative and procedural information that can act, guide, or modify the information flow in a cognitive system. In this respect, our model is conventional, following the analysis of Rumelhart and Ortony (1977) among others. The main claim of the model is that the information analysis and acquisition is guided by a control schema. Thus, if some decision has to be made, the requirements of the decision task act as requests for particular pieces of decision-relevant information. The set of these requests, plus their interrelationships and mutual constraints, is the control schema. In the domain of this research, the operation of the control schema is determined by a person's knowledge about potential attributes that describe stock quality and interrelationships among those attributes, knowledge about methods or strategies of stock market analysis, and skills used in acquiring decision relevant information, i.e., reading, skills for scanning information in tables, etc.

The control schema, we suggest, controls both the information acquisition task and the decision itself. Thus, in this task we have two pertinent behavioral end points, decision behavior and comprehension, with which to evaluate the control schema and its operations on input material (in the present case, text). We propose to investigate comprehension processes through the medium of recall tests, employing a model of text comprehension developed by Kintsch and van Dijk (1978) which permits us to predict recall as a function of the control schema employed in comprehension. Techniques for the analysis of decision processes are not presently as well developed.
We have developed a laboratory task which, while simplified, preserves the salient features of real life information analysis. We deal with short, stereotyped texts and a simple decision rule. Though the textual information by its very nature is fuzzy, the decision rule is deterministic. Specifically, subjects will be asked to act as stockbrokers, who have to decide whether to buy the stock of some fictitious company after reading the "Quarterly Report" of that company. These reports are brief paragraphs which contain carefully prescaled information on six categories relevant to the market activity of the stock. Subjects are aware of these categories and are trained to evaluate the texts in terms of the information in these six categories. The decision rule which they have to infer inductively is a simple conjunction of two of (the six) categories. If the information in a text is positive in both categories, the stock will always go up in price, indicating a "buy" decision. The control schema that the subject has to acquire to perform this task is, therefore, a simple one, consisting of requests for evaluative information in each of the six categories, plus a designation of two of the categories as decision-relevant.

We have thus moved the natural problem close to a traditional concept learning paradigm, with certain obvious disadvantages but the possibly compensating advantage that we can employ the powerful behavioral analyses developed for that paradigm. Thus, in addition to the decision and recall data, we have subjects state their hypotheses about decision-relevant categories on each trial, and we obtain their evaluations of
the six categories on part of the trials. One can look at this experiment as a concept learning study with texts as input, and hence with a fuzzy, or uncertain, dimensional and value structure.

Text construction. Real stockmarket reports contain quite unsystematic information. They tend to be skimpy, fragmentary, variable in form, and often inconsistent. They typically communicate information only about a few characteristics of the stock (say, capitalization or earnings), which will differ from one report to another about the same company. For experimental purposes, we had to construct artificial reports. The details of this procedure are described in Kozminsky, Bourne and Kintsch (Note 1).

Briefly, we first identified six categories of information relevant to stock behavior from actual stockmarket reports. A larger number could, of course, have been identified, but the following six categories are common, representative and readily defined:

1) General information -- information about economic conditions within this country and across the world which may have a bearing on the market in general, but does not have direct application to a specific company.

2) Capitalization -- information concerning the financial position of a specific company (assets, liabilities, cash on hand, credit status, existing loans, etc.).

3) Growth prospects and productivity -- information concerning past growth, near-term and long-term
expectations, possible mergers, expansions, and new products.

4) Sales -- historical information on company sales, near-term and long-term expectations, sales comparisons with other companies within the industry.

5) Earnings and profitability -- past earnings, near-term and long-term expectations and comparisons with other companies.

6) Dividends -- past and anticipated payments to stockholders.

Our next step toward the construction of messages was to select from real stock market information individual sentences which seemed to us to fall clearly into one or another of these six categories. We were able to find many such sentences. Often a good sentence would contain information pertinent to two or more categories, in which case the sentence was modified so as to address mainly one.

Information contained in these sentences ranged from extremely positive, for example, "Dividends will be doubled in the next fiscal year," to extremely negative, for example, "Sales have struck an all time low in the first quarter."

Through judicious selection and modification of the available sentences and several rating studies, we were able to develop a set of 20 sentences within each category which provided a uniform distribution of sentence values on a 5-point scale from extremely negative to extremely positive sentences.
These sentences were then combined into paragraphs containing information (positive or negative) on each of the six categories regarding a given company. Lists of six sentences, one of each category, were randomly selected and given to subjects who were asked to reorder the sentences in each list into the most rational (comprehensible, text-like) sequence. After reordering, subjects were asked to insert semantic connectives at their own discretion so that the ordered sentence lists made the best sense possible to them. A set of twenty texts, each providing information on every one of the six categories described above, was produced by this procedure. An example of such a text is given in Table 1. The first sentence contains information identifiable with the Dividends category. The second sentence is identified with capitalization, and so on. Notice that the sentences are not always unambiguous or explicit: For instance, the first sentence also refers to the financial position of the company—capitalization. In the second sentence the category identity is implicit. Thus, identifying the appropriate categories and evaluating them is a nontrivial task.

Insert Table 1 about here

Method

Subjects. Thirty-six subjects, 18 males and 18 females between the ages of 18 and 35 years participated in the experiment for five sessions each. They were paid $5.00 for their time, plus a bonus based on the quality of their performance in the experiment.
Materials. Subjects read 20 short texts each of about 100 words in length containing information relevant to six aspects of a fictitious company: General Factors, Capitalization, Growth, Sales, Earnings and Dividends. Each category of information in the text was pre-evaluated on a 5-point scale from 1 (negative) to 5 (positive information; see Kozminsky, Bourne, & Kintsch, Note 1). Unknown to the subject, two categories, Growth and Earnings, were chosen to be relevant to Buy/Not Buy decisions. The two relevant categories were combined according to a conjunctive rule to determine the proper decision on each trial. If the information for the two relevant categories in a text was evaluated 4 or more on a 5-point scale, the text was a positive example associated with a "Buy-stock" decision in anticipation that the stock price will go up. Ten positive texts (Buy) and ten negative texts (not Buy) were constructed. Three additional texts were prepared for use in a practice period.

Apparatus. The experiment was controlled by the Sigma 3 computer of the Computer Laboratory for Instruction in Psychological Research in the Department of Psychology, University of Colorado. The experimental program was written using the Display Terminal Experiment System (Spear, Note 2). The instructions and material was presented on a Four-Phase system CRT and responses were made on a five button panel interfaced with the computer.

Procedure. Experimental sessions were spaced over five consecutive days. The first session was a 90-minute training
session, and it was followed by four decision making sessions of approximately 30 minutes each. During the first training session, a subject went through the following sequence:

(a) The general nature of the task was explained. The subject was instructed to act as a stock broker recommending to a client to buy or not to buy a particular stock. In order to make a recommendation, the subject read a quarterly report about a company that issued the stock and made a recommendation according to the information in the report. For each correct recommendation the subject was paid 50 cents.

(b) A detailed description of the six informational categories was given.

(c) Then 18 sentences from the practice texts were displayed one at a time and the subject learned, with feedback, to assign sentences to their corresponding categories.

(d) In a second pass with the same 18 sentences, the subject evaluated the information in each sentence on the 5-point scale again with feedback.

(e) The nature of the decision rule, a conjunction, was explained and demonstrated with the three practice texts.

(f) Finally, a background description of the company in question was presented and followed by a short comprehension test.

The 20 texts were divided into four blocks, 5 texts in a block. In each of the four decision sessions a subject read five reports. Each subject was assigned to one of four
groups in which the text blocks were sequenced in counterbalanced design.

The decision sequence for each text was as follows:

(a) The subject selected two categories for a hypothesis about the relevant categories and marked them on a decision record. The decision record was in front of the subject throughout the experiment and he or she was allowed to examine it ad lib.

(b) The subject read the text and made a buy/not buy decision according to his or her expectation that a stock price will go up or down.

(c) Confidence in the decision was rated on a 5-point scale.

(d) Feedback about the decision was given and the subject recorded the outcome, right or wrong, on the decision record.

(e) The subject reinspected the text in order to decide whether and how to modify the hypothesis.

Subjects were told that they would, on occasion, be asked to recall the text just read. In instructions prior to the decision sessions, subjects were encouraged to recall as much as they could, not necessarily verbatim, and were told that for satisfactory recall they would get an additional bonus of $2.50 at the end of the experiment. (All subjects received this bonus.) These recall probes were given twice in each session immediately after the subject made a decision. Recall was not time limited. After recall, the subject evaluated the six informational categories in the text on a 5-point scale.
Then the sequence continued as in no recall trials. In each session the first trial and one other at random contained a recall test, a total of eight recall tests.

After the last decision session the subject received two final tasks, first, assigning sentences from four texts, to categories and, second, evaluating the informational content of these sentences. Also, a questionnaire was administered about the subjects' background, their attitude toward the experiment, and a verbal description of their decision behavior.

Results

There are three distinct sources of data in this experiment: (a) the hypothesis selection and decision data which were obtained on all trials of the experiment, (b) free recall protocols from some of the experimental trials, and (c) questionnaire data after the end of the experiment proper. We shall discuss these three data sources separately, emphasizing their interrelationships whenever possible. The decision and hypothesis selection results pertain to such questions as: How well did subject learn? What differences characterize the behavior of those subjects who did versus those who did not find a solution? The recall data are of interest because they permit one to infer what sort of information subjects had available when they made their decisions. Finally, the questionnaire data allow us to distinguish between problem solving strategies and to relate these strategies to characteristics of the decision processes as well as to the nature of the subjects' recall.
Hypothesis and decisions

Throughout this presentation we will distinguish between two subject groups: Learners are those subjects who, on some trial before the end of the experiment, selected the relevant categories Growth and Earnings as their hypothesis and maintained that hypothesis until the end of the experiment; The remaining subjects are grouped as Nonlearners. This criterion was used in preference to one based on decision performance because the uncertainty in the stimuli (texts) insured some proportion of decision error. On this criterion, 19 subjects were identified as learners and 17 as nonlearners. The mean of the first post-learning trial was 10.79 (approximately at the beginning of the third session), SD = 6.29. The distributions of trial numbers on which a relevant category was identified and maintained are shown in Figure 1.

-----------------------------
Insert Figure 1 about here
-----------------------------

When the subjects were asked, after the end of the experiment, to rate each category on a 10-point scale for its relevance to the decision task, learners rated the relevant categories higher than irrelevant categories by 6.55 points, while nonlearners produced only a 2.53 point difference between relevant and irrelevant categories. This difference between learners and nonlearners is statistically reliable, $t(34)=4.71$. Thus, the questionnaire results support the distinction between learners and nonlearners on the basis of intra-experimental hypothesis statements.
Decision performance. The proportions of correct decisions for learners and nonlearners over trials are shown in Figure 2. Average nonlearner performance was 53% correct, not reliably different from chance. Average learner performance was significantly better than chance, \( t(18) = 3.42 \). An analysis of variance with the factors learners vs. nonlearners, presentation order of the texts, and sessions yielded a significant F value for presentation order, \( F(3, 84) = 7.38 \), and a marginally significant interaction between learners and nonlearners over sessions, \( F(3, 84) = 2.42, p < .10 \). When the confidence ratings for the decisions were included in the analysis, this interaction reached a conventional significance level, \( F(3, 84) = 3.38 \).

-------------------------
Insert figure 2 about here
-------------------------

A comparison of pre- and postsolution decision performance for learners with the decisions made by nonlearners during the first and second half of the experimental sessions is presented in Table 2. While learners improve after they have decided upon the right hypothesis, nonlearners remain at a chance level. This interaction is highly significant statistically, \( F(1, 34) = 10.90 \).

-------------------------
Insert Table 2 about here
-------------------------

The evidence, then, is fairly clear: those 19 subjects classified as learners did indeed learn something about the appropriate decision rule, while the 17 nonlearners did not. It is important to note that this does not reflect an inability
of the nonlearners to understand the textual materials. After
the end of the experiment-proper, all subjects were asked to
classify and rate the sentences from four experimental texts.
Learners and nonlearners agreed equally well as to whether the
sentences were positive or negative (82% versus 80% correct,
$t(35) = .70$), and learners were only slightly more accurate
in identifying the category to which each sentence belonged
(92% correct versus 86%, $t(35) = 1.84$). Thus, the groups were
equivalent in their potential to learn. What, then, did the
learners do differently from the nonlearners?

Hypothesis selection. Subjects were requested to record
at the beginning of each trial their hypothesis about the
two relevant categories. The decisions they made were in
general consistent with their stated hypotheses, 92% of the
trials for learners, and 88% of the trials for the nonlearners,
the difference being statistically unreliable, $t(35) = .77$.
However, there were learner vs. nonlearner differences in a
number of other statistics that have been used in concept
formation experiments employing nonverbal stimuli and deter-
ministic rules (e.g., Millward & Spoehr, 1973). Some of the
more pertinent hypothesis selection characteristics are
summarized in Table 3 for presolution trials.

Insert Table 3 about here

(a) The likelihood of hypothesis change. Learners and
nonlearners did not differ in their hypothesis change rate. For
both groups the change rate dropped from about one hypothesis
change every two trials in the first half of presolution trials to about one change every three trials in the second half, F(1,30)=15.90. Relative to problems with simple structure this rate is low. We suspect that subjects in this study adjusted their hypothesis change rate by taking into account the uncertainty about properly categorizing and evaluating the information in the text.

(b) Category changes per hypothesis change. This measure decreased significantly from the first to the second half of the presolution trials, F(1,29)=11.20. There was however, a significant interaction, F(1,29)=4.40; when changing their hypothesis, learners consider more categories than nonlearners during the first half presolution trials.

(c) Proportion of incorrect decisions in a hypothesis run. Learners tolerated a ratio of about two incorrect decisions to every one correct before changing their hypothesis. For nonlearners this ratio was lower in the first half trials but not significantly so, F(1,30)=2.44. This measure is complementary to (a) and indicates the degree of adjustment to the uncertainty in the task. It also indicates tolerance to errors that can occur upon incorrect category perception.

(d) The likelihood of changing a hypothesis after an error. On the average, subjects changed their hypothesis only on 79% of the trials on which they had been told that their Buy/No Buy decision was incorrect, probably reflecting the fact that they were somewhat uncertain as to whether they had interpreted the information in the text correctly. However, the subjects' readiness to change after an error increased
sharply during the presolution trials, $F(1,30)=8.46$, with learners always being more likely to change than nonlearners, $F(1,30)=4.20$. Thus, learners were more successful in eliminating false hypotheses.

(e) Hypothesis recurrences. Learners were significantly less likely to try out a hypothesis again that had already been contraindicated, $F(1,27)=4.24$.

(f) Trials before a hypothesis is reconsidered. The better memory of learners for what they had done before is apparent from the fact that they waited significantly longer than nonlearners before they reconsidered an already-try hypothesis, $F(1,24)=8.56$.

In summary, these analyses highlight the following differences between learners and nonlearners. First, learners adopted a more global approach to this problem initially, considering more categories of information as potentially relevant during early trials (reports). Both groups adjusted their hypothesis change rate to the uncertainty in the task, but learners were more systematic and more conservative in the actual changes made. Furthermore, learners were more sensitive to the implications of feedback regarding their decisions, making more appropriate changes and only on occasions dictated by feedback. Finally, learners were less likely to retest hypotheses eliminated by information provided on earlier trials; which implies more confidence in previous decisions, better memory, or a more effective problem solving strategy. This summary bears remarkable similarity to conclusions about
good and bad problem solvers in vastly different and simpler problem domains (e.g., Bourne, 1965; Millward & Spoehr, 1973). Our results may, therefore, be taken to suggest the existence of much the same basic processes underlying a wide range of comprehension-abstraction-decision behaviors.

Recall

In every one of the four sessions of the experiment subjects were asked to recall two of the texts immediately after they read them and made their decisions. In each session the first text and one other randomly selected text were recalled. Each one of the eight recalled texts was propositionalized using the method developed by Kintsch (1974). Then, each recall protocol was scored by template matching it to the propositional test base (see Turner and Greene, 1978 for details of this procedure). As a further measure of comprehension, reading times were obtained for all texts.

There were no differences between learners (33.90% recall) and nonlearners (34.26% recall) in amount recalled (i.e., number of propositions). In an analysis of variance with the factors learners-nonlearners, sessions, first- and second-recalled text in each session, and presentation blocks, only the main effects of session, first-vs-second recall test within sessions, and blocks were significant. Recall improved over the four experimental sessions, from 27.52% in the first session, to 34.50%, 36.81%, and 37.51% in the following sessions, $F(3,84)=6.49$. The first text in each session yielded significantly
higher scores than the second text, 36.45% vs. 31.72%, $F(1,28) = 15.88$. Blocks were also significant statistically, $F(3,84) = 10.06$, indicating a lack of homogeneity among the texts used in the experiment.

Further, learners and nonlearners did not differ in reading rate, where a corresponding analysis yielded only two significant interactions, between first-and-second test and sessions, and first-and-second test and blocks, $F(3,84) = 4.06$ and $F(3,84) = 12.38$, respectively. These merely indicate the expected variability among texts in the various presentation blocks, but also question the usefulness of the second test in each session on which reading rates first decreased and then increased over sessions. The difference of primary interest, between learners and nonlearners, is simply not present. Indeed, reading rates are almost identical: 3.13 seconds per proposition for learners, and 3.15 seconds per proposition for nonlearners. (Seconds per propositions rather than total reading times were used to control for the differences in the lengths among the experimental texts.)

There are, however, differences between learners and nonlearners if one takes into account the nature of their recall, rather than merely the overall amount. Learners remember those aspects of the texts that are relevant to the decision task, while nonlearners do not show a comparable selectivity and appear more text-bound than goal-determined. Suppose one asks whether, for each of the six task dimensions, a subject recalled enough information to permit a correct evaluation on that dimension
(i.e., whether the dimension is positive or negative). Since the various dimensions differ greatly in their saliency depending upon the particular text that happens to be used, the scores from the first time each text was used (when the subjects as yet knew nothing about the decision task) were subtracted from all scores. Thus, evaluative memory improvement scores are obtained. On the average, these were higher for learners (.47) than for nonlearners (.02), F(1,31)=4.12, p=.052. Similarly, when they were asked, at the end of each recall trial, to rate the six critical task dimensions as to their positive or negative informational values, learners showed considerable improvement during the course of the experiment, from 66% correct on presolution trials to 79% after they had found the correct solution. Nonlearners, on the other hand, gave only 63% correct evaluations over the entire experimental task. (Note learners achieved 82% correct and nonlearners 80% when rating the same sentences outside the decision task context).

The selectivity of recall by learners is most clearly demonstrated in Figure 3, which plots percent recall for the relevant categories Growth and Earnings. Again, improvement over the first trial recall is shown rather than raw scores in order to control for saliency effects. The data are averaged by aligning scores, such that for each subject the trial on which a correct hypothesis was formed corresponds to the first postsolution trial. Subjects who never learned contribute only to the presolution portion of the learning curve.
Figure 3 shows that as learning proceeded, subjects recalled increasingly more from the relevant dimensions, and significantly more once they had the correct solution. (The final drop in the curve is based on a small number of scores and is therefore difficult to interpret.) The gradual improvement on presolution trials and the high level of recall of the relevant dimensions thereafter might be an artifact of the general increase in recall over sessions. Figure 4 shows, however, that this improvement was mainly due to the better recall of the relevant information by those subjects who learned the proper decision rule, confirming the interpretation of Figure 3 as a selection process on relevant text information.

Insert figures 3 and 4 about here

Since total recall was the same for learners and non-learners, while learners recalled more relevant information than nonlearners, the latter must have recalled relatively more task-irrelevant information from the experimental texts. This is particularly true for one type of easily identifiable irrelevant information, namely, the sentence connectives that were inserted into the experimental paragraphs to make them appear more like real texts than lists of sentences. For connective recall (improvement over the first time a text was read in order to control for differential saliency effects), the scores for nonlearners are four times as high as those for learners, 16.65% and 3.97% respectively, $F(1,31)=3.48$.

A recall model. Learners retain relevant information, while nonlearners show little selectivity. A rather simple
explanation of this difference follows from the model for text comprehension and recall proposed by Kintsch and van Dijk (1978). According to that model, people remember what they do. Reading the experimental paragraphs leaves certain kinds of memory traces. Evaluation of six task categories has its own further memorial consequences. All subjects undertake these activities, but only the learners consistently select the categories of Growth and Earnings as relevant and base their decisions upon their evaluation of these categories. This fact, we propose, leads to the recall enhancement on postsolution trials shown in Figure 3.

Theoretically, we assume that the basic local processes in reading (that is, the microprocesses) are the same for all subjects, but that there are two kinds of macroprocesses. One is common to all subjects and is controlled by the instruction to rate and evaluate all six task dimensions within each text. This process leads to the selection of evaluation-relevant information for each category. In this task, then, the "gist" of each paragraph consists of the six category evaluations, and the information used in arriving at these evaluations is presumed to have a special status in memory (as macropropositions). But, in addition, for learners only, higher level macropropositions can be identified on postsolution trials, namely the decision-relevant evaluations. Thus, as far as memory is concerned, we need to distinguish between micropropositions, and two kinds of macropropositions, those for the relevant and those for the irrelevant categories.
Consider how this model fits the recall data in the present experiment. Because we are interested in studying the effects of mastering the decision task on recall, we shall examine first the data from Sessions 3 and 4, when at least some subjects were making correct decisions. Only the first recall in each session was used for model fitting, because the second recall was significantly lower and might lead to distortion in the parameter estimates. Thus, there were four different texts to work with. For each of these, Growth and Earnings were the relevant categories. For these two categories a subject's data were included in the analysis only if these categories were indeed selected (hypothesized) as relevant by that subject. For the other four categories, a complementary selection was made: a subject's data in any one of these categories were considered only if the category was not the subject's hypothesis on that trial. Thus, not all data were used in the model analysis: for the relevant categories, we looked at only those subjects who thought these categories in fact were relevant, and for the irrelevant categories we took only those subjects who thought them irrelevant. Under these circumstances, unambiguous theoretical predictions were possible. However, we had only 314 protocols points to work with (70 were deleted for the reasons explained), which is not sufficient to support sophisticated statistical estimation procedures.

Detailed descriptions of the model are available in Kintsch and van Dijk (1978). Here, we shall merely sketch the main features of the model by working through a small example.
The model accepts as input a proposition list that represents the semantic context of the text. This list is obtained through handcoding the actual text according to a set of semi-explicit rules. The model attempts to simulate the organizational processes that this semantic input undergoes during comprehension, distinguishing their local (microprocesses) and global (macro-processes) components. The model reads a text (or, rather, the corresponding proposition list) in chunks, tries to determine the coherence of the information in each chunk, and relates the various chunks via common referents carried over from one chunk to the next in a limited capacity buffer. Recall predictions follow from this model because each time a proposition is processed, it may be stored in memory (and later retrieved on the recall test) with some probability $p$, to be estimated from the data. Each proposition is processed at least once upon input, but some propositions are processed more frequently because they are held over from one processing cycle to the next in the short-term buffer.

The macroprocesses occur under the control of the decision schema. The purpose of reading these texts is to identify and to evaluate the six informational categories. Hence, all propositions that are directly relevant to this operation receive further processing and are stored in memory with some probability $m$, another parameter of the model. Finally, the two categories upon which the actual decision is based are processed once more, the memorial consequences of which is represented by a third model parameter, $n$. 
As an example, Table 4 shows how the fourth sentence from the text from Table 1 is processed by the model. First of all, since it is longer than 21 words, the sentence is broken into two chunks for processing. Propositions P37, P39, and P42 are selected as macrostructure-relevant, because they are needed to infer the identity of the category--Earnings--and whether the information about Earnings in this text is positive or negative. Since Earnings, furthermore, is a category upon which a decision is actually made, these propositions are processed twice at the macrolevel, indicated by the operators $M$ and $H$ in Table 4. Most macropropositions are selected on the basis of a key-word approach in the present texts (e.g. "Earnings", "Dividends"); where this approach fails, normal conditions or consequences for any of the six text categories are determined from which evaluative statements can be inferred. The macro-processes also provide a superordinate proposition for the more local processing of the text, in this case P39. The microprocesses then attempt to construct a coherent text base around this superordinate, by connecting all propositions to it in a tree-like structure as shown in Table 4. Connections are made whenever two propositions share a common referent; when this condition does not obtain, bridging links must be inferred. Once a text base is constructed, there are rules in the theory to select a few propositions to be held over for the next processing cycle so that a single coherent structure is generated rather than a separate tree for each processing cycle. The
propositions to be held over are selected on the basis of their superordinate position in the text base as well as their recency. In the present example, P37 and P39 are held over, and the input propositions from the next cycle are directly annexed to them, resulting in a coherent representation for our sample sentence. The memorial effects of these microprocesses are indicated by the operator S in Table 4 (e.g. the S-operator is applied twice to P37 because that proposition was processed in both processing cycles).

In the model, the S, M and N operators are simple probabilistic equations with three parameters, p, m and n. These can be estimated by fitting the actual pattern of recall to the theoretical predictions. In the present case, a version of the method of moments was used for estimating these parameters, and other parameters of the model, such as the input size which we set at 21 words, or the buffer capacity, s=2, were simply guesses that appeared reasonable from other experience with this model. (As noted, more data would be needed for the use of more sophisticated statistical procedures.) However, even with these nonoptimal estimates, the resulting recall predictions appear to mirror reasonably well the main trends in the data in Table 4. Indeed, over all four texts, the predictions are excellent. Table 5 shows the p, m and n parameters that were estimated for the four texts, the goodness of fit of the model to the data in terms of a chi-square criterion, and the correlation between observed and predicted recall patterns. The average correlation of .84 is comparable with other
applications of the model (e.g. Kintsch & van Dijk, 1978; Spilich, Chiesi, Vesonder, & Voss, 1979). Note that for two of the paragraphs, our estimate of \( n \), which represents the effects of actually making a decision on the two relevant categories Growth and Earnings, turned out to be zero; 20 learning trials were apparently not sufficient to produce strong selection effects in all of the texts.

The good fit of the model to the recall data is important because it provides converging evidence for the theoretical claim that mastering the decision task involves the acquisition of a control schema. In its crudest form, this schema consists of six independent requests for evaluative information about each of the experimental categories, plus a (conjunctive) decision rule. What we have shown now is that this same schema also controls the subjects' recall. We still need to investigate the relationship between what is remembered and the particular hypotheses that a subject selects for the decision task, especially with respect to strategy differences. Do learners somehow use their memory differently than nonlearners?

-----------------------------
Insert Table 5 about here
-----------------------------

Memory and hypothesis selection. Given that a subject changed his hypothesis on Trial \( i \) and that he recalled a text on that trial, each category state was determined relative to the subject's hypothesis on Trial \( i-1 \). There are four possible category state transitions: Kept. A category was in the subject's hypothesis on Trial \( i-1 \) and it was maintained in the subject's hypothesis on Trial \( i \). Nonselected. A category
was not in subjects hypothesis on Trial i-1 and it was not selected to be in the hypothesis set on Trial i. Selected. A category was not in the hypothesis set on Trial i-1 and it entered the set on Trial i. Dropped. A category was in subject's hypothesis on Trial i-1 and it was dropped from the set on Trial i.

As an example, suppose that the hypothesis set on Trial i-1 consisted of the categories Capitalization and Dividends and that on Trial i the set was composed of Capitalization and Earnings. In this case, the category Capitalization was kept from Trial i-1 to Trial i; the categories General Factors, Growth, and Sales were nonselected; Earnings was selected; and, Dividends dropped from the hypothesis set on Trial i. Category state change can be similarly defined for the Trial i to Trial i+1 transition. Average category state transition recall was computed as a function of the category state transition from Trial i-1 to Trial i and as a function of the state transition from Trial i to Trial i+1.

The relationship between category recall and subjects hypothesis changes on presolution trials is shown in Figure 5. Note, that the same recall data were used to obtain the two transition patterns. The percent category recall enhancement scores were obtained as the difference between the category recall on Trial i to the average category recall on Session 1 to factor out text saliency effect. These scores were computed separately for learners and nonlearners.

------------------------
Insert Figure 5 about here
------------------------
An analysis of variance was performed on recall enhancement scores using three factors: Subject type (learners and non-learners), hypothesis state transitions, and trials transitions (Trial $i-1$ to Trial $i$ and Trial $i$ to Trial $i+1$). This analysis indicated two interactions: Subjects by hypothesis state transitions, $F(3,90)=5.16$, and a hypothesis state transitions by trial transitions, $F(3,90)=3.56$. The first interaction primarily reflects a recall advantage in the Kept and Nonselected states for learners. The second demonstrates a recall increase for the Kept and Dropped state in Trial $i$ to Trial $i+1$ transition.

There are several interesting differences between learners and nonlearners in Figure 5. Consistent with hypothesis data, we show in a different way that the set monitored includes categories that are not yet selected to be in the hypothesis set, as well as those already in it, as evident from their superior recall of the Kept and Nonselected state categories compared to nonlearners. Nonlearners recall is less constructive, reflecting their concern with the last category selection operation and the categories about to be dropped from the hypothesis set. This could imply that nonlearners use less effective hypothesis selection and testing strategies than learners. Nonlearners invest more resources in the operations of selecting and dropping hypotheses on each trial, resulting in better memory for those categories because of the extra processing. As we shall see, the questionnaire data support such an explanation.
Strategy Analysis

Successful performance in any concept identification task relies on the utilization of an appropriate strategy. Since Bruner, Goodnow, and Austin (1956), considerable evidence has accumulated on the appropriate strategy for problems which are simpler but isomorphic to the underlying structure of the task used in this experiment and on the relationship between strategy and overall performance (Laughlin, 1973). Despite the complexities of the task confronting subjects in this experiment, it is appropriate to search for the strategies subjects used and to determine whether they played any role in the distinction between learners and nonlearners.

Instructions to subjects encouraged purposeful, conscious, testing of hypothesis about the categories relevant to Buy/not Buy decisions. The task requires the subject to analyze each text into separate categories of information and to evaluate at least those categories which are pertinent to the current basis of decision (hypothesis). If subjects adopt some detectable strategy, analytic or otherwise, it would have implications not only for speed of learning, but also for the nature of subjects search through categories of information within each text and for the subjects' recall of text.

One way to detect the subject's strategy is to examine the relationships that exist from trial to trial among hypothesis statements. We preferred, however, to use a strategy measure as independent of the primary task as possible. Therefore, we administered a questionnaire to all subjects at the end of the
experiment designed to reveal what, if any, strategy was used. The responses of subjects on this questionnaire indicated, rather clearly, four types of strategies. This classification system is not exactly like any other reported in the literature, but contains within it strategies which have been identified by Levine (1975), among others, in related tasks. We label these strategies as follows. First, a Global strategy similar to Bruner, Goodnow and Austin's Wholistic strategy or what Levine called Global focusing or Perfect processing. There are four subjects who qualify as this type. These are subjects who, from the outset, considered all six categories of information potentially relevant. They kept track of values in each category over successive texts until a category was identified as irrelevant. They attempted logically to eliminate categories on the basis of inconsistencies between category values and changes in the stock's price. Second, a strategy adopted by 22 subjects which we call Focusing. These subjects exhibit a pattern of performance similar to Bruner, Goodnow and Austin's Partist strategy. Their technique is to evaluate only two categories of information in each text, the two they had (sometimes arbitrarily) selected for their hypothesis. They are guided primarily by information in the text just preceding the current trial. They were locally consistent in their hypothesis, but often made inconsistent hypothesis selections relative to information given earlier in the trial sequence.

Each of these two strategies is highly analytic in its orientation. The subject enters each trial with some hypothesis,
reads the text in the light of that hypothesis, assigns information to categories contained within the hypothesis, evaluates those categories, makes a decision on the basis of the evaluation, and changes the hypothesis in the light of feedback received.

The remaining subjects were distinctly different in their orientation toward the task. We refer to one group of 8 subjects as Intuitive. Their decisions tended to be influenced by the overall atmosphere of the text more than by their own hypotheses. They reported difficulty in categorizing the textual information and in dealing with the uncertainty in the task. As a consequence they tended to rely on an overall impression of the paragraphs, positive or negative, as a basis of decision. Their hypotheses were formulated independently of feedback. The final group of 2 subjects is referred to as Unsystematic. They reported being overwhelmed by the task. They did not attempt to categorize textual information and therefore performed no evaluation of category information. Their responses were often described as guesses.

This strategy analysis was made entirely on the basis of the questionnaire data. However, the system of classes that results has implications for performance within the decision task itself, and a series of analyses was undertaken to examine these implications. An obvious prediction holds for the learning data. As noted above, the task requires an analytical orientation. We therefore, expect subjects who adopt an analytic strategy to perform significantly better than subjects who adopt an
intuitive strategy. For purposes of this analysis, we group Global and Focusing subjects (n=26) as analytic in orientation, and Intuitive plus Unsystematic (n=10) as intuitive in orientation. The groups were further divided as to whether they met the criterion of learning. Seventeen analytic subjects but only 2 intuitive subjects met the criterion, resulting in a significant \( \chi^2(1) = 5.97 \). Thus, clearly, subjects who reported following an analytic strategy throughout the learning sessions were more likely to perform successfully on the task itself.

Perhaps the most important implications of the strategy analysis pertain to recall. We examined the following possibility. Subjects should perform well on recall of a given category of information to the extent that they process that category deeply during reading. Subjects will process to the depth of available resources any category of information which is presently a part of their stated hypothesis or otherwise under consideration as a potential determiner of the stock price. If the subject is focusing during presolution trials on certain categories of information to the exclusion of others, recall across categories will have high variability. If the subject considers all categories equally viable, variation in category recall will be low. Because they process two categories of information deeply and the remainder on a shallow level at best, subjects who focus should exhibit high variability in category recall. Subjects in the Intuitive group should show relatively low variability. A critical aspect of our prediction is the Global group. These subjects begin by considering all categories as equally relevant. Therefore, category recall should be low in variability on early
trials. But these subjects also narrow, by a process of elimination from six categories to the relevant two. At that point, recall should be as variable as that exhibited by the Focus group. We expect, then, that category variability and recall will increase significantly over presolution trials for Global subjects. It is difficult to know how unsystematic subjects might perform and our predictions exclude this group.

Figure 6 shows the standard deviation of category recall for the Global, Focus, and Intuitive groups during Vincentized halves of presolution trials. These standard deviations are expressed as difference scores, correcting for variability in Session 1. All groups show an increase in variability, attributable to the increase in overall recall from first to second half of trials. Focus subjects have significantly higher variability than Intuitive subjects. In contrast, Global subjects performed during the first half more like Intuitive subjects and during the second half more like Focus subjects. The statistical test reflecting this interaction is significant, $F(1,26)=7.06$.

------------------
Insert Figure 6 about here
------------------

Tentative and fragmentary though they may be, these data suggest the need for two distinct models to account for problem solving-decision making performance. One reflects a deliberate, hypothesis-testing approach to the extraction of information from text and the other reflects a more passive approach driven by the overall atmosphere created by a report (e.g., Hammond, Note 4).
The following is a sketch of a model for analytic subjects (Model I). The subject enters each trial and reads each passage with a control schema which accepts information pertinent to \( c \) categories. For subjects whom we have referred to as Focussers, \( c=2 \); for subjects who fall into the Global category, \( c \) is the total number of categories, here six. In principle, however, subjects may record information about any number of categories. Upon presentation of the text passage, the subject reads to identify information pertinent to the \( c \) categories of the schema. There is uncertainty at this point, however, for we know that the sentences are not perfectly categorizable. Once the information has been categorized, subjects process further the information in each of their critical categories for evaluation. On the basis of category evaluation, the subject makes a decision (Buy/not Buy) and receives feedback. Feedback is processed in such a way as to correct the current control schema for the next trial. Feedback confirms the viability of some categories and infirms others. Confirmed categories remain a part of the schema for the next trial. Infirmed categories are deleted. This process cycles over successive passages until sufficient information is accumulated to override uncertainty in the category and evaluation processes and to bring the subject to the accurate identification of the two relevant categories of information.

A different model is required for those subjects who used an impoverished control schema in this task (Model II), the intuitive subjects (Hammond, Note 4). These subjects enter a
trial without a hypothesis. They bypass the information categorization stage, moving directly to evaluation. Their evaluation reflects the overall tone of what has been read rather than the value of each sentence, ignoring the implications of feedback for their decision.

There are a number of implications of these two models. Model II subjects might show serial order effects both in recall and decision performance. The value of early and (especially) late categories in a given text ought to have more impact on overall text evaluation and therefore on decision. Model I subjects should show little or no serial order effects on decisions to the extent that $c$ is small. Further, Model II subjects should read and make decisions more rapidly because they require fewer processing steps to arrive at a decision point. Finally, the hypothesis statements of Model II subjects should be far less affected by feedback from preceding trials than hypotheses given by Model I subjects. There are probably other implications of these models which will appear as we work them out in detail. But the data of the present experiment are too few to make assessments at the present time.

General Discussion

This research is exploratory. Our purpose was to provide a framework, however sketchy, which would be useful as a general model for information analytic behaviors in complex, naturalistic environments. Our goal was to outline a model, anchored in empirical evidence, which would provide a basis for further, theory-guided research. Such a model requires, at a minimum,
integration of more limited theories of text comprehension, abstraction, and decision making.

The present experiment makes some progress toward this goal. The data show, among other things, that there is a sensible interface between the text comprehension ideas of Kintsch and van Dijk (1978) and the well-worked-out notions of hypothesis behavior (Levine, 1975), which have heretofore been limited in their application to rather simple and artificial conceptual tasks. Indeed, one of the most striking outcomes of the present research is data to establish the greater generality of hypothesis testing notions.

The outstanding characteristic of all of analyses was the great variability among individuals. Despite common instructions and common task requirements, our subjects differed widely in the strategies they used to identify or abstract the relevant categories of information from the texts. While the task itself requires a subject to deal analytically with the text and the majority of subjects did adopt an analytic strategy, a considerable range of approaches from near optimal to entirely intuitive was observed. Furthermore, differences in abstraction strategy correlated with differences in recall performance. Where the subject's strategy called for category processing, e.g. hypothesis testing the relevance of a given category, recall was high. Categories excluded from subject's hypothesis, and therefore presumably not processed as deeply in that particular text were recalled relatively poorly. As another example, consider the high recall of sentence connectives by
nonlearners. This indicates that nonlearners were using a different control schema in reading these texts than the learners, one that probably was not very well defined, but that was more determined by the properties of the text itself than by the decision task. The latter, as we have seen, controlled the reading of the learners. As far as the model of Kintsch and van Dijk (1978) is concerned, the nature of the processes employed by the learners and nonlearners in this task may very well have been the same, but the outcome was very different because of the different control schemata used by these two groups of subjects.

We can conclude at this juncture that individual differences must be part of any general model of information processing. To the extent that these differences can be traced through a variety of performance components, comprehension-recall-abstraction-decision, they become critical in the development of any theoretical framework for understanding information-analytic behavior.

We suspect that the eventual framework, whatever the details, will be schema based. If one accepts this idea, then it becomes possible to investigate some additional interesting empirical questions. There are many ways to learn in this task, that is, to transform the initial task-schema, established by the experimental instructions, into the desired decision schema. What are the most effective ways? Which hypothesis selection and testing strategies succeed, and which fail? What sort of
information does the learner remember from reading the texts, and what do the nonlearners remember? How does text processing and memory relate to the concept learning strategies subjects employ? Tentative answers to these questions have been suggested above, but we plan further experiments to obtain more conclusive evidence.

Eventually, we hope that the kind of results we have obtained will provide a conceptual basis for the development of training procedures in information analysis tasks. Before that is possible, however, we need to investigate longer texts that provide the kind of informational overload that characterizes natural information-analytic tasks. We need to investigate redundant and contradictory texts requiring more complex control schemata and inferential processes. Finally, we need to go beyond the restrictions imposed by a simple deterministic decision rule. It is not at all clear that in more complex tasks the same strategies will be optimal. But we have here, at the very least, developed a method with which we can find out.
Footnote

This research was supported by a contract from the Office of Naval Research, N00014-78-C-0433. This report is publication Number 89 from the Institute of Intellectual Studies at the University of Colorado. Requests for reprints should be sent to Ely Kozminsky, Department of Psychology, University of Colorado, Boulder, Colorado, 80309.

1. A significance level of .05 is assumed hereafter.

2. The microprocesses have been completely formalized in terms of a computer program in Miller and Kintsch (Note 3). No such formalization exists as yet for the other parts of the theory; the application of the macro-operators is intuitive, though constrained by the principles presented in Kintsch and van Dijk (1978).
Reference Notes

1. Kozminsky, E., Bourne, L. E., Jr. & Kintsch, W.
   Comprehension and analysis of information in text: I.
   Construction and evaluation of brief texts. Technical Report
   No. 82-ONR, Institute for the Study of Intellectual Behavior,

2. Spear, T. L. Display Terminal Experiment System User's
   Manual. CLIPR Publication No. 12, Computer Laboratory for
   Instruction in Psychological Research, Department of Psychology,

3. Miller, J. R. & Kintsch, W. Readability and recall of short
   prose passages: A theoretical analysis. Unpublished manu-
   script, University of Colorado, 1979.

4. Hammond, K. R. The Unification of Psychological Decision
   Theory and Social Judgment Theory. Unpublished manuscript,
   Human Judgment Center, Institute of Behavioral Science,
   University of Colorado, 1979.
References


Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F.

Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 1979, 18, 275-290.*

Table 1

Sample paragraph used in the experiment

The company has skipped the dividend again this year advancing cash flow problems as the cause. Furthermore, banks have refused to renew the company's credit line without representation on the Board of Directors. However, recent strengthening in the monthly composite of leading indicators provides an appearance of a better underlying tone to the economy and company sales could reach $420-440 million, up 25% from the last fiscal year. But, considering the higher prospective shipment costs, earnings can fall in the range of $6.00-$7.00 per share next year rather than the previously estimated $7.00-$8.00. Thus, we anticipate a period of slower growth next year between 3-4% per annum.
Table 2
Proportion correct decisions of learners as a function of solution state and of nonlearners as a function of sessions

<table>
<thead>
<tr>
<th></th>
<th>Before solution</th>
<th>After Solution</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td>.46</td>
<td>.73</td>
<td>.60</td>
</tr>
<tr>
<td>First two sessions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last two sessions</td>
<td>.53</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>Nonlearners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.50</td>
<td>.63</td>
<td>.56</td>
</tr>
</tbody>
</table>
Table 3
Presolution hypotheses selection characteristics of learners and nonlearners

(a) Probability of hypothesis change per trial.  
(b) Average number of category changes per hypothesis change.

<table>
<thead>
<tr>
<th></th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td>.59</td>
<td>.35</td>
<td>.47</td>
<td>1.40</td>
<td>1.08</td>
<td>1.24</td>
</tr>
<tr>
<td>Nonlearners</td>
<td>.55</td>
<td>.32</td>
<td>.44</td>
<td>1.21</td>
<td>1.13</td>
<td>1.17</td>
</tr>
<tr>
<td>Mean</td>
<td>.57</td>
<td>.34</td>
<td>.46</td>
<td>1.30</td>
<td>1.11</td>
<td>1.20</td>
</tr>
</tbody>
</table>

(c) Proportion of incorrect decisions in a hypothesis run.  
(d) Probability of changing a hypothesis after an error.

<table>
<thead>
<tr>
<th></th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td>.65</td>
<td>.62</td>
<td>.64</td>
<td>.74</td>
<td>.97</td>
<td>.86</td>
</tr>
<tr>
<td>Nonlearners</td>
<td>.49</td>
<td>.67</td>
<td>.58</td>
<td>.60</td>
<td>.83</td>
<td>.72</td>
</tr>
<tr>
<td>Mean</td>
<td>.57</td>
<td>.65</td>
<td>.61</td>
<td>.67</td>
<td>.90</td>
<td>.79</td>
</tr>
</tbody>
</table>

(e) Proportion of hypothesis recurrences to number of hypotheses changes.  
(f) Average trials distance between hypotheses repetition.

<table>
<thead>
<tr>
<th></th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
<th>First Half</th>
<th>Second Half</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td>.18</td>
<td>.47</td>
<td>.32</td>
<td>4.10</td>
<td>8.20</td>
<td>6.15</td>
</tr>
<tr>
<td>Nonlearners</td>
<td>.39</td>
<td>.54</td>
<td>.46</td>
<td>3.55</td>
<td>5.42</td>
<td>4.48</td>
</tr>
<tr>
<td>Mean</td>
<td>.28</td>
<td>.50</td>
<td>.39</td>
<td>3.82</td>
<td>6.81</td>
<td>5.32</td>
</tr>
</tbody>
</table>
Table 4

The processing of a sample sentence from the text example in Table 1

<table>
<thead>
<tr>
<th>SENTENCE</th>
<th>THEORY</th>
<th>PREDICTED</th>
<th>OBSERVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 (BUT, P32)</td>
<td>S</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>32 (CONSIDERING, P33, P36)</td>
<td>S</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>33 (HIGHER, COST)</td>
<td>S</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>34 (SHIPMENT, COST)</td>
<td>S</td>
<td>16</td>
<td>35</td>
</tr>
<tr>
<td>35 (PROSPECTIVE, COST)</td>
<td>S</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>36 (CAN, P37)</td>
<td>S</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>37 (RANGE, EARNINGS, $6-7PSH)</td>
<td>SSMN</td>
<td>79</td>
<td>83</td>
</tr>
<tr>
<td>38 (NEXT-YEAR, P36)</td>
<td>S</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>39 (RATHER-THAN, P37, P42)</td>
<td>SSMN</td>
<td>82</td>
<td>83</td>
</tr>
</tbody>
</table>

BREAK AFTER 21 WORDS

<table>
<thead>
<tr>
<th>SENTENCE</th>
<th>THEORY</th>
<th>PREDICTED</th>
<th>OBSERVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 (PREVIOUSLY, P41)</td>
<td>S</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>41 (ESTIMATED, P42)</td>
<td>S</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>42 (RANGE, EARNINGS, $7-8PSH)</td>
<td>SSMN</td>
<td>79</td>
<td>83</td>
</tr>
</tbody>
</table>

SENTENCE

Processing Cycle Text Base:

<table>
<thead>
<tr>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6</td>
</tr>
</tbody>
</table>

Cycle i 39 36 38 33 34 37

Cycle i+1 39 37 42 40 41

MACRO PROCESSES:

Cue-word "EARNINGS"
Selects P37, P42;
Add P39 as connective
Table 5

Goodness of fit of the model
for the four experimental texts

Summary for S-2:

<table>
<thead>
<tr>
<th>Text</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
<th>m</th>
<th>n</th>
<th>r</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46.51</td>
<td>36</td>
<td>.30</td>
<td>.86</td>
<td>.68</td>
<td>.81</td>
<td>.66</td>
</tr>
<tr>
<td>B</td>
<td>18.54</td>
<td>20</td>
<td>.22</td>
<td>.44</td>
<td>.00</td>
<td>.84</td>
<td>.71</td>
</tr>
<tr>
<td>C</td>
<td>22.75</td>
<td>21</td>
<td>.27</td>
<td>.80</td>
<td>.00</td>
<td>.87</td>
<td>.76</td>
</tr>
<tr>
<td>D</td>
<td>27.94</td>
<td>19</td>
<td>.16</td>
<td>.70</td>
<td>.31</td>
<td>.83</td>
<td>.69</td>
</tr>
</tbody>
</table>

|      | 115.74   | 96 | .24 | .70 | .25 | .84 | .71   |
Figure Captions

1. The proportion of subjects who identified correctly the categories Growth (full line), Earnings (broken line), and both Growth and Earnings (dotted line) on the 20 experimental problems and the final test.

2. The proportion of correct decisions for learners (full line) and nonlearners (broken line) for the four experimental sessions.

3. Improvement in recall of the relevant sentences on pre- and post-solution trails. The dotted line indicates the 95% confidence interval.

4. Improvement in recall over the four experimental sessions for irrelevant sentences (broken line, open circles), relevant sentences-unlearned (broken line - full circles), and relevant sentences-learned (full line).

5. The relationship between category recall and hypothesis changes on presolution trials for learners (shaded bars) and non-learners (unshaded bars).

6. The variability in category recall on presolution trials for Global (closed circles), Focus (open circles) and Intuitive (squares) subjects.
<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Address/Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. E. Aiken</td>
<td>Navy Personnel Research &amp; Development Center</td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Dr. Jack R. Bersting</td>
<td>Provost &amp; Academic Dean</td>
<td>U. S. Naval Postgraduate School</td>
</tr>
<tr>
<td>Dr. Robert Breaux</td>
<td>Code 11-71</td>
<td>NAVTRADEPUCN</td>
</tr>
<tr>
<td>Dr. Charles E. Davis</td>
<td>GIN Branch Office 53E S. Clark Street</td>
<td>Chicago, IL 60605</td>
</tr>
<tr>
<td>Mr. James S. Duve</td>
<td>Chief, Human Factors Laboratory</td>
<td>Naval Training Equipment Center</td>
</tr>
<tr>
<td>Dr. Richard Elster</td>
<td>Dept. of Administrative Sciences</td>
<td>Naval Postgraduate School</td>
</tr>
<tr>
<td>Dr. Pat Federico</td>
<td>Navy Personnel R &amp; D Center</td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Dr. John Ferguson, MSc, USN</td>
<td>Naval Medical R&amp;D Command (Code 44)</td>
<td>National Naval Medical Center</td>
</tr>
<tr>
<td>Dr. John Ford</td>
<td>Navy Personnel R &amp; D Center</td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Dr. Eugene Gloye</td>
<td>GIN Branch Office 1030 E. Green Street</td>
<td>Pasadena, CA 91101</td>
</tr>
<tr>
<td>Dr. Steve Harris</td>
<td>Code L522</td>
<td>NAVD1</td>
</tr>
<tr>
<td>LCDR Charles W. Hutchins</td>
<td>Naval Air Systems Command</td>
<td>444 Jefferson Place #1</td>
</tr>
<tr>
<td>Dr. Norman J. Kerr</td>
<td>Chief of Naval Technical Training</td>
<td>Naval Air Station Memphis</td>
</tr>
<tr>
<td>Dr. Leonard Kromer</td>
<td>Navy Personnel R&amp;D Center</td>
<td>San Diego, CA 95152</td>
</tr>
<tr>
<td>Mr. J. Andrew surgeons</td>
<td>Chairman, Leadership &amp; Law Dept.</td>
<td>Div. of Professional Development</td>
</tr>
<tr>
<td>Mr. James Lester</td>
<td>GIN Branch Office</td>
<td>Bldg. 114, Sec. D 666 Summer Street</td>
</tr>
<tr>
<td>Dr. William L. Maloy</td>
<td>Principal Civilian Advisor for Education &amp; Training Naval Training Command</td>
<td>Code 00A</td>
</tr>
<tr>
<td>Capt. Richard L. Martin</td>
<td>USS Francis Marion (LPA-249)</td>
<td>FPO New York, NY 09901</td>
</tr>
<tr>
<td>Dr. James McGraw</td>
<td>Navy Personnel R&amp;D Center</td>
<td>Code 306</td>
</tr>
<tr>
<td>Dr. William Montague</td>
<td>LCDC</td>
<td>University of Pittsburgh</td>
</tr>
<tr>
<td>Commander Officer III</td>
<td>U.S. Naval Amphibious School</td>
<td>Coronado, CA 92155</td>
</tr>
<tr>
<td>Commander Officer IV</td>
<td>Naval Health Research Center</td>
<td>Attn: Library</td>
</tr>
<tr>
<td>Commander Officer V</td>
<td>Naval Medical R&amp;D Command</td>
<td>Code 44</td>
</tr>
<tr>
<td>Librarian</td>
<td>Navy Personnel R&amp;D Center</td>
<td>San Diego, CA 95152</td>
</tr>
<tr>
<td>Commander Officer VIII</td>
<td>Naval Research Laboratory</td>
<td>Code 262</td>
</tr>
<tr>
<td>Dr. John Olsen</td>
<td>Chief of Naval Education &amp; Training Support</td>
<td>USS Francis Marion (LPA-249)</td>
</tr>
<tr>
<td>Mr. Arnold Rubenstein</td>
<td>Naval Personnel Support Technology</td>
<td>Naval Material Command (OBM-244)</td>
</tr>
<tr>
<td>Dr. Richard Pollack</td>
<td>Academic Computing Center</td>
<td>U.S. Naval Academy</td>
</tr>
<tr>
<td>Dr. Worth Scanland</td>
<td>Chief of Naval Education &amp; Training Support</td>
<td>USS Francis Marion (LPA-249)</td>
</tr>
<tr>
<td>Tech. Support, Code 201</td>
<td>Navy Personnel R&amp;D Center</td>
<td>San Diego, CA 95152</td>
</tr>
<tr>
<td>Personnel &amp; Training Research Programs</td>
<td>(Code 305)</td>
<td>Office of Naval Research Arlington, VA 22217</td>
</tr>
<tr>
<td>Psychological Officer</td>
<td>Office of Naval Research Branch</td>
<td>223 Old Maryland Road</td>
</tr>
<tr>
<td>Psychological Officer</td>
<td>GIN Branch Office 1030 East Green Street</td>
<td>Pasadena, CA 91101</td>
</tr>
<tr>
<td>Scientific Director</td>
<td>Office of Naval Research</td>
<td>Scientific Liaison Group/Tokyo American Embassy APO San Francisco, CA 96503</td>
</tr>
<tr>
<td>Head, Research, Development &amp; Studies (OP012X) Office of the Chief of Naval Operations Washington, DC 20370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific Advisor to the Chief of Naval Personnel (PEARS-OR) Naval Bureau of Personnel Room 4410, Arlington Annex Washington, DC 20370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr. Arnold Rubenstein</td>
<td>Naval Personnel Support Technology</td>
<td>Naval Material Command (OBM-244)</td>
</tr>
<tr>
<td>Dr. Richard Pollack</td>
<td>Academic Computing Center</td>
<td>U.S. Naval Academy</td>
</tr>
<tr>
<td>Dr. Worth Scanland</td>
<td>Chief of Naval Education &amp; Training Support</td>
<td>USS Francis Marion (LPA-249)</td>
</tr>
<tr>
<td>A. A. Stahl</td>
<td>Tech. Support, Code 201</td>
<td>Navy Personnel R&amp;D Center</td>
</tr>
</tbody>
</table>