Information Choice and Utilization in an Experiment on Default Prediction

A. RASHAD ABDEL-KHALIK* AND KAMAL M. EL-SHESHA†

In an effort to encourage progress in the area of human information processing, Einhorn [1976] suggested that researchers in this area begin to incorporate information search and choice in their experimental designs. Traditionally, experiments in this area have presented subjects with a predetermined set of cues to utilize in forming certain judgments. This type of design, Einhorn observes, can reduce the involvement of subjects in the experimental task. In fact, few studies have replicated the richness of the information search environment, and few have implicitly studied information choice and acquisition in experimentation (Slovic, Fischhoff, and Lichtenstein [1977, pp. 7-9]). The only known studies that explicitly employed information choice by subjects are those by Pankoff and Virgil [1970] and Payne [1976].

In this paper, we report the design and results of an experiment in human judgment in which decision makers were allowed to choose the information cues they used in making their judgments. The results of this study show that the subjects' choice of information, rather than their processing of the chosen cues, was the limiting factor in predicting the environmental event—in this case, default on debt. While the approach adopted here does not completely overcome the problems mentioned by Einhorn, it does provide a step in that direction.

* Professor, University of Florida; † Associate Professor, Georgia State University. We wish to thank Bipin B. Ajinkya, Bill Messier, Doug Snowball, Dick Chesley, and especially the reviewer for providing useful and constructive criticism. We also express appreciation to William E. Pierce and Hindupur V. Ramakrishna for assisting in data collection and analysis. This project was partially supported by the Accounting Research Center of the University of Florida and by the Research Council of the College of Business of Georgia State University. [Accepted for publication November 1979.]
The rest of this paper is divided into several sections: (a) a brief review of relevant literature; (b) statement of the problem and the objectives of the study; (c) a brief discussion of how we adapted the lens model approach to include information acquisition by subjects; (d) the design of the study and the type of subjects who acted as decision makers; (e) the results in terms of comparison with several performance indices; (f) replication of the experiment and attempts to deal with the incompatibility of priors; and (g) concluding remarks and limitations.

1. Prediction of Failure and Information Processing

Research concerning the prediction of failure has followed two approaches: (1) using mathematical or statistical models based upon publicly available data (see Ohlson [1980] for a review of recent work), and (2) the ability of human decision makers to use accounting information in discriminating between failed and nonfailed pairs of firms. As summarized in Bernstein [1974, pp. 462-65], earlier studies concerning the prediction of failure using publicly available accounting information have utilized a mixed mode of time-series (trends) and cross-sectional data. The recent work by Beaver [1966], Altman [1968], and Deakin [1972] used cross-sectional analysis only by using a paired sample of failed (mostly bankrupt) and nonfailed firms. Recently, Ohlson [1980] also used accounting information in a cross-sectional analysis to form a probabilistic prediction of bankruptcy.

Some of the work which utilized the second approach was characterized by the stimulus-black box response paradigm (e.g., Abdel-khalik [1973]), while other work emphasized the methods that judges apply in integrating, differentiating, and processing information (Libby [1975; 1976] and Kennedy [1975]).

2. Research Problem and Objective

In reviewing this prior evidence, two issues of interest emerge. The first relates to the definition of failure used in prior studies (typically bankruptcy). To illustrate the importance of this issue, we will refer to Beaver's graphs [1966, p. 82] depicting the temporal behavior of the three ratios he found most predictive of failure. These graphs show that the cash flow/total debt and net income/total assets ratios exhibited a continuous decline in the health of the failed firms (mostly bankrupt) throughout the five years prior to the failure. The third ratio, total debt/total assets, depicted a high level of debt-to-equity structure of failed firms relative to

---

1 It should be noted that Beaver [1966] defined bankruptcy as "the inability of a firm to pay its financial obligations as they mature"; however, his data included fifty-nine bankrupt firms, sixteen involved nonpayment of preferred dividends, three bond defaults, and one overdraft. Most authors adopt the same definition (except Altman [1968]), but they use mostly bankrupt firms.
that of nonfailed firms for all five years, with a sharply increasing trend in this ratio up to bankruptcy. Hence, financial ratios are structurally at their worst (but most predictive) level in the year prior to bankruptcy. Indeed, Ohlson [1980] found that annual reports of many bankrupt firms were actually released after they declared bankruptcy.

Lending analysts, however, may be just as interested (or even more interested) in predicting "default on loans," which, in most cases, is an event that occurs earlier than a year away from bankruptcy. Since more firms default on debt than go bankrupt, we used the event of default on debt as the environmental event to be predicted.

The second problem with earlier studies is that they provided a predetermined set of information cues for judges to utilize in making decisions. Although subjects could make selective use of cues given to them, under this approach the initial information choice set was determined by the researcher alone. Alternatively, if decision makers could share in the information choice at the initial stage, and also continue to exercise additional choice by selectively using only some of the acquired cues, their task involvement would more likely increase. While the selective use of cues probably takes place in all experiments, having subjects participate in making the initial choice of information cues does constitute an additional experimental structure that has not been studied previously. As Einhorn pointed out, in forming judgments in a real-life decision situation, "information must be searched for—it is not given" [1976, p. 200]. He adds: "As I have tried to point out, in most real judgment situations, the judge takes a much more active role by searching for information, forming hypotheses, etc." [1976, p. 205]. In his view, highly structured experiments (with no sharing of information choice) could lead judges to take a passive role in performing the task.

While the objectives of this study are threefold—(a) to involve the decision maker in the judgment task under consideration, (b) to improve the definition of the environmental event designated as a failure, and (c) to provide indices for possible evaluation of the quality of choice and use of information by judges—its main one is to evaluate the relative contribution of each of the two acts made by loan officers in predicting default—information choice and information processing.

Indeed we know of only two studies which dealt with this issue. In the study by Pankoff and Virgil [1970], financial analysts in the St. Louis area participated in an experiment which involved acquiring information, forecasting stock prices, and making portfolio decisions. Pankoff and Virgil simulated a setting where each analyst had an initial endowment, a limited amount of information and the possibility of acquiring additional information at a predetermined (hypothetical) cost. They observed that most frequently purchased items were earnings per share, company sales, industry sales, and stock price indices. They concluded that the quality of performance was not generally a function of acquired information, and that the usefulness of information acquisition appeared to be in satisfying the analysts' desire for inputs, which they called "input demand." In another study, Payne [1976] simulated information seeking as a search-for-an-apartment problem, where subjects were asked to seek information about each apartment from among a limited set of cues.
3. Research Design and Method

APPROACH

In adapting the experimental design of the lens model so as to incorporate acts of information choice and processing, we looked upon the problem of information choice as consisting of three stages: (i) Based on theoretical or empirical models, the researcher determines a set of information cues which are highly associated with and predictive of the environmental event. (ii) At the second stage, the decision maker or the judge is given the option of selecting a limited number of cues from the above set, which was initially determined by the researcher. (iii) The information cues chosen by the judge or the decision maker could still be selectively used in forming the judgment, which is an act of further exercising information choice.

Several considerations need to be discussed here. First, the judge or the decision maker could be either a human or a mathematical (a mechanical) model. The information cues chosen are uniquely determined by the judge making the choice. The components of the chosen cues (corresponding to the second stage above) must be constrained (either by cost or by number), otherwise human decision makers are likely to acquire all information cues determined by the researcher in the first stage. Similarly, the "shopping list" of possible cues determined by the researcher at the first stage should not be too large because the measurement of a chosen cue must be made available on demand to the judge during the conduct of the experiment.

The choice processes exercised by the mathematical model (or the mechanical decision maker) and by the human judge had similarities, but differed in the procedure implemented in this experiment to affect the choice. The model search was accomplished by having a mechanical model scan all measures of the cues in the shopping list (determined by the researcher), rank-order them by their predictive power (or according to some optimization rule), and list the set of cues that would maximize the prediction. For the human search, however, judges were given only the names (or descriptions) of all the cues included in the shopping list or the information choice set (not their measurements) and were permitted to acquire measurements of a subset of those cues according to their own preferences.

Given that the prediction of the environmental event is the principal task, three different sources of predictions are applicable: human predic-

\[^{3}\text{If subjects are not constrained in their choice, and if they perceive some of the information cues presented in the "shopping list" as irrelevant, their performance should not be expected to improve. The results of experimentation in the literature on multiple-cue probability learning (Castellan [1977]) show that (a) the larger the number of irrelevant cues, the worse the performance of judges, and (b) the larger the number of irrelevant cues, the slower the rate of learning [1977, pp. 36-37].}\]
tions made by the decision makers, predictions based on "models of man," and predictions made by a mathematical model independent of the others. Each type of prediction utilizes information cues chosen either by the mathematical model (the mechanical decision maker), or by human decision makers. The resulting six combinations of the information choice and prediction form the six strategies that are shown in table 1.

The predictive validity of each of these six strategies is assessed on the basis of percentage of correct classification (hit rates) as used by Libby [1975; 1976] instead of correlations. Other than changing the approach to include information choice, which increases the number of combinations, the three predictions (by human judges, by "their models," and by the environmental mathematical model) are fundamentally those highlighted by Goldberg [1970] and used by others. Measures of predictive validity of these six strategies are used below in obtaining indices of performance evaluation.

FIRMS IN THE SAMPLE

The Index of Corporate Events in issues of the Disclosure Journal (1973-75) was screened to obtain a list of firms that defaulted for nonpayment of debt, but not for other reasons such as covenant violation. Two constraining criteria reduced the sample size: (a) the availability of the firm's financial statements in publicly available sources such as Moody's for five years prior to default, and (b) the need to find a set of firms which had not defaulted on debt and which were comparable to the defaulted firms in terms of size (as measured either by total assets or by total sales) and industry classification. The resultant sample size consisted of sixteen defaulted firms which were individually matched with another sixteen firms representing twelve industries. These firms represented the experimental sample shown in panel A of table 2. Subsequently, the same procedure was repeated for the selection of another sample of defaulting firms from among those listed in the Index of Corporate

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategies of Predictions</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction of Environmental Event (Default)</th>
<th>Information Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>by Human</td>
<td>by Mathematical Models</td>
</tr>
<tr>
<td>By human</td>
<td>(1) **HP</td>
</tr>
<tr>
<td>By &quot;models of man&quot;</td>
<td>(2) **MP,</td>
</tr>
<tr>
<td>By environmental models (optimal)</td>
<td>(3) **MP,</td>
</tr>
<tr>
<td></td>
<td>(4) **HP</td>
</tr>
<tr>
<td></td>
<td>(5) **MP,</td>
</tr>
<tr>
<td></td>
<td>(6) **MP,</td>
</tr>
</tbody>
</table>

*H = human, M = model, S = selection or choice, P = prediction; the subscripts refer to the basis of generating the model as to predicting human response (s) or the environmental event (e).

**HP | HS** is the prediction of the environmental event by human subjects based on their own information selection. This is frequently called the achievement index.

**MP, | HS** is the prediction of the environmental event using the mechanical model built to structure human responses in a linear form using human-selected information.

**MP, | MS** is the model's environmental predictability using the model-selected information.
Events of 1975–76. After applying the same screening criteria discussed above, a sample of fourteen defaulted firms was obtained. This sample is shown in panel B of table 2. The relevance of the two panels is discussed later.

FINANCIAL RATIOS AND THEIR DISCRIMINATING POWER

Previous studies of the usefulness of financial ratios in predicting failure used both trends and point estimates of the ratios, with the more contemporary ones using point estimates only. Since multicollinearity between financial ratios leads to a disagreement about the identity of most of the discriminating financial ratios, we simply compiled an information set (call it the shopping list) consisting of ten financial ratios that were used in previous studies and eight trends. A trend was estimated by the average annual change in a financial item over the five years preceding the year in which default occurred. A step-wise discriminant analysis was used for model search; two models were estimated, one for each of the panels A and B. The results of the models’ information choice and prediction are shown in table 2.

In order to validate these results, the discriminant analysis model produced by panel B was applied to the data of panel A. Given that both panels have the same nondefault (control) sample, the importance of this cross-validation lies in the classification of the defaulted sample of panel A, which correctly predicted 57 percent, since the entire nondefault

---

**Table 2**

<table>
<thead>
<tr>
<th>Samples Used in the Study</th>
<th>Control Sample 16 Nondefaulted Firms</th>
<th>Experimental Sample 16 Defaulted Firms</th>
<th>Validation Sample 14 Defaulted Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of correct classification of own sample</td>
<td>90.6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Information used by the mechanical model</td>
<td>Total debt/Total assets</td>
<td>Current ratio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend of L.T.D./N.W.</td>
<td>Quick assets/Sales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Net income/T.A.</td>
<td>Net income/Sales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend of N.I./Sales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend of N.I./T.A.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

1. The nondefault firms were used as if they were a control sample to be compared with an experimental sample. It is true that even though the firms in panel B were not matched (as were those in panel A), the results obtained for panel B were even better than those obtained for panel A. Such a result did not lead us to reject the notion of the "control" sample, although, in retrospect, it would appear that selecting another sample would have simplified the presentation.

2. The recent study by Ohlson [1980] suggested four factors: (1) a measure of size, (2) a measure of financial structure, (3) a measure(s) of performance, and (4) a measure(s) of liquidity. We have used all except size.
AN EXPERIMENT ON DEFAULT PREDICTION

This additional step was employed in order to evaluate the bounds of correct classification obtained by using mechanical or mathematical models. As shown in table 2, the upper bound for the experimental sample is 0.906 correct classification, whereas the lower bound is obtained from the cross-validation of the defaulted firms, which is about 0.57.

SUBJECTS AND THE TASK

The decision makers who participated in this study consisted of twenty-eight commercial lending officers in a large metropolitan area. All twenty-eight subjects were contacted through a senior administrator in their respective banks, and there was no experimental mortality of subject participation. The average lending experience of the participating officers was about eight years. A capable doctoral student (who had some business experience in management) was recruited to conduct the interviews with these lending officers. Each interview was prearranged and was conducted individually in the office of each analyst.

In the instructions given to the analysts, the objective of the study was stated as the need to compare the relative importance of various financial indicators in forming judgments about whether a firm will default on the payment of its debt within one year or two years from the date on which the information was prepared. They were informed that (1) the firms were real; (2) some of them had actually defaulted on the payment of debt at a date subsequent to the time period covered by the financial information included in the shopping list; (3) the firms were presented to them in a random order; and (4) the proportion of defaulted to nondefaulted firms was not representative of the true proportion observed in real life. They were also informed that all the information cues named in the information choice set were derived from audited financial statements through the year before the default date.

The task consisted of making judgments about whether each of the firms in the list should be classified in the defaulted or the nondefaulted category. The judgment was to be based on the information acquired under a given set of assumptions. There were two sequences of such acts of information acquisition, use, and judgment. The assumptions given to them were stated as follows: (1) Your (hypothetical) budget for purchasing information for the purpose of predicting default is 100 points. (2) Information items may be supplied to you at (hypothetical) cost. The price schedule is provided below. (3) You are permitted to purchase no more than four pieces of information at any one time. (4) If you purchase one piece of data, it will be provided to you for all firms.

We used points instead of dollars out of concern for the potential impact of the differential utility for money. The choice of four information
items as a maximum per each cycle of information acquisition-use-judgment was basically arbitrary. In conducting the experiment, analysts were encouraged to use at least one information cue in each of the cycles.

The experimental materials consisted of two single-spaced pages of instruction, the first containing the firm’s identification number (from 1 to 32) and its industry and a space for responding to questions in the first round of the information purchase-use-decision. The second page was used for making judgments in the second round. A single sheet of paper was also provided to each lending officer listing the names of the eighteen information cues constituting the choice (or shopping) list. In each interview, the research assistant started with a general conversation for about five minutes, during which he inquired about the officer’s lending experience and other issues that might appear to be of interest. The assistant would then hand the written instructions to the lending officer and ask him to read it and follow instructions. The last item on the instruction sheet reminded the officer that he had to reveal his information choice, at which time the research assistant would provide sheets containing only the information requested and charge the officer’s account with the cost (points) stated in the supply table. In order to facilitate this process, each of the eighteen information items was produced on a single detached column and presented for each of the thirty-two firms exactly in the same sequence and spacing as the listing of the firms on the answer sheet. So, if an officer wanted to buy information items numbered 2 (net income/sales) and 4 (cash flow/total debt), he would get two narrow sheets of paper so that he could look at both information items in the sequence in which each of the thirty-two firms was presented.

After making their information choice in the first round, lending officers were told how many points remained in their budget. They then would examine the information and make judgments on each of the thirty-two firms as “default,” “no default.” A similar process was repeated for the second cycle of buy-use-decide. When the task was completed, almost all of the subjects indicated that they thought the experiment was more engaging than they had initially expected, and they wished to know how “well” they did. The only feedback given to the subjects between the two rounds was that they could improve their judgments by selecting more information. At the end of each round of judgment, lending officers were requested to rank the importance of each item they purchased as they perceived its contribution toward facilitating their judgments.

**MOTIVATION OF DECISION MAKERS**

A measure of performance inherent in the task was provided to the decision makers who participated in this study. In the instructions about the experiment, they were told:

> Since we know which firms in the sample have actually defaulted on the payment of
AN EXPERIMENT ON DEFAULT PREDICTION

their debt in the fiscal year following the date of the information provided here, we can derive a measure of the quality of utilization and the information to predict the default and to be able to compare your effectiveness with other lending officers. The measure of such a quantity is computed as:

\[
\frac{\text{Index of Value of Information}}{\text{The Unused Balance of Information Budget}} \times \frac{\text{Number of Correct Predictions}}{\text{Number of Total Predictions}}
\]

This measure of performance could be maximized by minimizing purchases of information acquisition, while at the same time increasing the number of correct predictions. Naturally, we did not know what behavioral effects would be induced by telling subjects that their performance would be compared with other lending officers, except that they were enthused about the comparison. Although incorporating an implicit performance evaluation function could have provided additional motivation to participants in the experiment, we make no claims that this is a perfect surrogate for a real-life reward. The motivation provided is constrained by the nature of the experiment and the findings have to be so interpreted.

4. Analysis and Results

SUBJECTS' INFORMATION CHOICE PROFILE

Interviews with the lending analysts were arranged over a period of two months. An interview lasted, on the average, about seventy-five minutes, which were devoted to reading the instructional material, contemplating information acquisition, making judgments, acquiring additional information, and revising those judgments. The experimenter was always present to supply the information on request and to collect the results. The frequency count of information items acquired by the officers during the two rounds is shown on the left-hand side of table 3. As shown, the most frequently purchased pieces of information and those perceived as most useful were earnings trend, current ratio, cash flow to total debt, and trend of cash flow to total debt. The judges were permitted to repeat the cycle in order to improve their judgments; the additional information items most frequently acquired were total debt to total assets, long-term debt to net worth, and trend of net income to sales.

Note that there is one-to-one correspondence between the relative frequency with which an information item was acquired and the perceived importance attached to it by the analysts. Furthermore, we observed that those analysts who acquired more information cues in the first round continued to acquire relatively more in the second round.

7 The concept of cost as a weight in the percentage of correct classification was used only as a constraint on the choice of information. Granted, the constraint has no external validity, but the concern here was more of internal validity and motivation of subjects.
\begin{table}
\centering
\caption{Frequencies of Information Choice by Lending Officers (Made by Task Execution) and Their Own Ranking of Importance (Made after Task Execution)}
\begin{tabular}{|l|c|c|c|c|c|c|c|c|}
\hline
\textbf{Information Variable} & \multicolumn{3}{c|}{\textbf{No. of Times* Selected in}} & \multicolumn{5}{c|}{\textbf{Frequency of Judges Own Rank of Perceived Importance}} \\
\textbf{No. and Name} & \textbf{1st Round} & \textbf{2d Round} & \textbf{Both Rounds} & \textbf{1st} & \textbf{2d} & \textbf{3d} & \textbf{4th} & \textbf{5th} & \textbf{6th} \\
\hline
1. Net income/Total assets & 0 & 1 & 1 & 1 & 3 & 2 & 1 & & \\
2. Net income/Sales & 7 & 0 & 7 & 2 & 1 & 4 & 2 & 4 & \\
3. Total debt/Total assets & 7 & 6 (1) & 13 (3) & 4 & 3 & 2 & 1 & 2 & \\
4. Cash flow/Total debt & (3.5) 10 & 2 & 12 & 6 & 2 & 3 & 2 & & \\
5. Long-term debt/Net worth & 5 & 5 (2.5) & 10 & 1 & 2 & 3 & 3 & 1 & \\
6. Current assets/Current liab. & (2) 11 & 2 & 13 (3) & 6 & 2 & 3 & 2 & & \\
7. Quick assets/Sales & 0 & 1 & 1 & & & & & & \\
8. Quick assets/Current liab. & 8 & 4 (4) & 12 & & & & & & \\
9. Working capital/Sales & 3 & 1 & 4 & & & & & & \\
10. Cash at year-end/Total debt & 0 & 0 & 0 & & & & & & \\
11. Earnings trend & (1) 16 & 2 & 18 (1) & 9 & 2 & 4 & 2 & & 1 \\
12. Sales trend & 2 & 3 & 5 & 1 & 1 & 1 & 2 & & \\
13. Current ratio trend & 8 & 3 & 11 & 2 & 5 & 2 & 2 & & \\
15. Trend of W.C./Sales & 5 & 2 & 7 & 1 & 2 & 3 & 1 & & \\
17. Trend of N.I./Sales & 3 & 5 (2.5) & 8 & 1 & 2 & 1 & 2 & 2 & \\
18. Trend of cash flow/T.D. & (3.5) 10 & 3 & 13 (3) & 4 & 4 & 2 & 2 & & 1 \\
\hline
\end{tabular}
\end{table}

* Numbers in parentheses represent the ranking of most frequently used information.
ADDITIONAL BENCHMARKS

In addition to the measures of performance to be derived from the three prediction strategies in table 1, additional standards or benchmarks for evaluation of human performance were developed. These were:

**BM1. Using Beaver’s Ratios:** The three ratios that Beaver found most predictive of failure (cash flow/total debt, net income/total assets, and total debt/total assets) were utilized in a multivariate discriminant analysis to classify the sample of thirty-two firms into defaulted and nondefaulted. Although Beaver’s study focused on predicting bankruptcy instead of default, application of the three ratios here resulted in a 0.72 correct classification. This benchmark is quite useful for our purposes because (a) it is the product of variables generated by a sample quite independent of the samples used here, (b) Beaver’s three ratios were part of the information choice set available to lending officers, and (c) two of these ratios were ranked as fourth and sixth on the list of most frequently acquired information cues.

**BM2. Using Most Frequently Acquired Cues:** Similar to the development of the first benchmark, the four variables most frequently acquired by lending analysts in this experiment were employed in a discriminant analysis to evaluate their predictive ability. The resulting classification yielded about 0.75 correct predictions.

**BM3. Using the Model Generated by Experimental Sample:** As shown in table 2, the results of this benchmark (denoted MP_{eMS}, in table 1) were about 0.906 correct classification using the six information items selected by the (environmental) mathematical model shown in table 2. Because the model estimation was based on the data of the experimental sample (which was also the sample used by lending officers), its correct classification is actually an upper bound for the performance expected of human subjects.

**BM4. Using the Model Generated by the Validation Sample:** In this case, a model was generated by the sample displayed in panel B of table 2 and was subsequently projected into the experimental sample. The difference between this case and the third benchmark lies in the fact that the model was based on information contained in a different sample. The model estimation using the data of panel B resulted in 100-percent correct classification of both defaulted and nondefaulted firms; and when it was projected into the experimental sample it misclassified six of the default firms as nondefault. Due to the commonality of the control (nondefault) group between panels A and B, this measure derives a benchmark only for the classification of defaulted firms (0.57 correct classification), which is based on cross-validation. (If total number of correct classification is used, default and nondefault, the percentage increases to 0.78.)

* Using the most frequently acquired six cues in a discriminant analysis did not change that classification.
5. Evaluation of Human Performance

JUDGMENT REVISIONS

After selecting no more than four items of information and using them in making predictions, subjects were permitted to select up to four more information items and then revise their judgments. The second round of information choice, use, and judgment was suggested only if the experimenter noted misclassification, which was true for all the subjects. As a result of the revisions, the accuracy of subjects' predictions improved on average by $0.007$. The probability that this difference was significant is at least $0.18$. The benefits of additional information in this case varied, in that eight subjects performed better by using the additional information, seven did worse, and thirteen showed no change in the accuracy of their predictions from the first round. Consequently, all the measures that follow are based on the predictive results of the second round.

INDICES OF INFORMATION CHOICE AND PROCESSING

The quality of predictions obtained by the six strategies (see table 1) was used to represent different measurements of validity. The results based on percentage of correct prediction (hit rates) are shown in table 4. These results assume that the loss functions of the misclassification are symmetric for the default and nondefault predictions.

Using these results, the following comparisons are possible:

(i) The extent to which the choice of information cues by human

<table>
<thead>
<tr>
<th>Strategy #</th>
<th>Denoted</th>
<th>Evaluating the Predictive Ability of the Environmental Event</th>
<th>% of Correct Prediction (Hit Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>HP</td>
<td>HS</td>
<td>The judge's validity (achievement index) using human-selected information.</td>
</tr>
<tr>
<td>S2*</td>
<td>MP,</td>
<td>HS</td>
<td>The validity of the &quot;model of man&quot; using human-selected information.</td>
</tr>
<tr>
<td>S3</td>
<td>MP,</td>
<td>HS</td>
<td>Environmental mathematical model using information selected by humans.</td>
</tr>
<tr>
<td>S4</td>
<td>not studied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>not studied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>MP,</td>
<td>MS</td>
<td>Environmental mathematical model using model-selected information.</td>
</tr>
</tbody>
</table>

* Strategy S2 uses the linear model estimated from human responses and the information cues that they used to predict the environmental event. The linear model estimated from human response (the model of man) correctly predicted 0.84 of the responses of the judges. The 0.625 correct classification shown in the table is the percentage of correct prediction of the environmental event using the "model of man."
decision makers deviates from optimal, reflected in the difference between
the quality of prediction obtained by strategies (S3) and (S6). Although
both (S3) and (S6) utilize a mathematical model of a similar basic
structure to predict the environmental event, one model uses information
cues selected by the mathematical model, whereas the other uses the
information cues chosen by human judges. The index, as measured by
\((MP_e|HS) - (MP_e|MS)\), is equal to -0.23, which suggests that human
judges appear to have made a suboptimal choice of the cues they utilized.

(ii) The extent to which information processing by humans deviates
from optimal processing, evaluated by the difference between strategies
(S1) and (S3). While each uses the information cues selected by human
judges to predict the environmental event, they differ on the processor of
those cues—humans in strategy (S1) and models in strategy (S3). This
difference, measured by \((HP_e|HS) - (MP_e|HS)\), is equal to -0.05,
which suggests a minor advantage in processing by the mathematical
model rather than human processing.

(iii) The extent to which weighting of cues by humans (assuming a
linear model) deviates from optimal, based on a comparison of the results
of strategies (S2) and (S3). This comparison shows that, on the average,
weighting of information cues by humans results in a performance level
slightly below (off by 0.05) the level of the model.

Based on the results of this study, it appears that human predictions
of default fell short of the predictive ability of the mathematical models
utilized here, mainly because of the less than optimal choice of informa-
tion cues. Neither the processing nor the weighting of cues appeared to
contribute significantly to this lower performance, although the model
slightly outperformed humans on that dimension as well.

### COMPARISON WITH OTHER BENCHMARKS

The quality of overall performance of human decision makers can be
evaluated further by comparison with the benchmarks. Given that the
mean responses of the officers in predicting the actual events was about
0.625, a test of significance could be accomplished by comparing the
percentage of correct classifications of each of the benchmarks against a
confidence interval about 0.625. The 0.90 confidence interval shows the
bounds to fall between 0.76 and 0.48 (see table 5), suggesting that loan
officers performed worse than either of the first two benchmarks in the
absolute, but not in the statistical sense and that they performed worse
than the third benchmark in both absolute and statistical terms.

Furthermore, the performance of loan officers in classifying defaulted
firms of the experimental sample averaged 0.42 correct classification,
which is below the 0.58 level achieved by the fourth benchmark. All of
these comparisons suggest that the loan officers generally performed
worse than any of the mechanical models used in generating the bench-
marks.
TABLE 5
Summary of Subjects Performance (Association Between HP, HS, Other Benchmarks, and Actual)

<table>
<thead>
<tr>
<th>Correct Prediction Ratio (CC)</th>
<th>Number of Judges</th>
<th>Relation to Benchmarks*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70 &lt; CC ≤ 0.75</td>
<td>2</td>
<td>S</td>
</tr>
<tr>
<td>0.66 ≤ CC &lt; 0.70</td>
<td>11</td>
<td>S</td>
</tr>
<tr>
<td>0.60 &lt; CC &lt; 0.66</td>
<td>10</td>
<td>W</td>
</tr>
<tr>
<td>CC &lt; 0.60</td>
<td>5</td>
<td>W</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>(S = Same; W = Worse)</td>
</tr>
</tbody>
</table>

Classification Matrix (Average Response)

<table>
<thead>
<tr>
<th>Predicted (HP, HS)</th>
<th>Actual Event</th>
<th>D</th>
<th>ND</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>6.75</td>
<td>9.25</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>ND</td>
<td>2.8</td>
<td>13.2</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9.55</td>
<td>22.45</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

Mean % of correct classification = 0.62
Phi = 0.28

Statistical Significance

H₀: Mean of correct classification by (HP | HS) is significantly different from correct classification of BMᵢ, i = 1, 2, 3

Test: a 0.90 confidence interval =

\[
\frac{0.62 \pm 1.65 \sqrt{0.62(0.38)}}{32} = 0.62 \pm 0.14
\]

Result: Correct classification of BM₁ and BM₂ fall within bounds, but BM₃ does not.

These are directional relations only. Results of comparison with BM₄ are included in the text.

6. Dealing with Priors

A confounding variable, however, was that the mechanical models were implemented under priors of equal size of default and no-default populations, while loan officers did not have these same priors. This problem was evaluated as follows. First, we replicated the study with an explicit statement about the composition of the sample as one-half or 0.50 defaulted and one-half or 0.50 no-default. Second, we introduced some naive, rule-of-thumb benchmarks which, in effect, are neutral with respect to the composition of our sample.

THE REPLICATION

Because of the expense and time associated with structured interviews, we randomly selected ten of the twenty-nine officers who performed the earlier study and asked them “to perform some more analysis for the study before providing them with the feedback.” Each of the ten agreed to grant our research assistant a thirty-minute interview. At the interviews, the research assistant indicated that the sample consisted of one-half defaulted and one-half no-default firms and that we would appreciate their performing the task one more time with this knowledge in mind. They were asked to perform only one round of information choice, use, and judgments, after which they were told of their performance on the earlier study. A different order might have been desirable, but this order of task-then-feedback was adopted in an attempt to avoid introducing any additional information, except the change in priors.
The replication was performed after a period of five to six months had elapsed since the first study. The average length of the replication interviews was about forty minutes. The results, reported in table 6, are somewhat discouraging (or encouraging, depending on one's viewpoint). As shown, the average performance of the ten officers was not generally better with the stated priors of 50 percent–50 percent, compared to the average performance without an explicit quantification of these priors. The mean change in the percentage of correct prediction was about 1 percent, and the standard deviation changed from 6.8 to 5.7 percent. Our research assistant observed that most of the subjects made judgments on the basis of the information they selected, then apparently balanced their judgments in order to achieve equal numbers in the categories of default and no-default. This balancing process along with the higher consensus appear to be the only meaningful outcomes of the explicit introduction of the percentage composition of the sample as 0.50 default and 0.50 no-default. Although this result was not consistent with our expectation, we felt no need to continue the replication with the rest of the subjects.

RULES OF THUMB AS PREDICTORS

Even if loan officers were not told about the composition of our sample, the application of some rules of thumb could have generated better predictions than their own utilization of the information they acquired. Observing that twenty-two of the twenty-eight subjects acquired either

<table>
<thead>
<tr>
<th>Officer Coded Number</th>
<th>% of Correct Classification of the Study without a Statement of Priors</th>
<th>% of Correct Classification of the Replication with a Statement of Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>66</td>
<td>56</td>
</tr>
<tr>
<td>23</td>
<td>53</td>
<td>63</td>
</tr>
<tr>
<td>1</td>
<td>62</td>
<td>56</td>
</tr>
<tr>
<td>22</td>
<td>50</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>69</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>63</td>
<td>56</td>
</tr>
<tr>
<td>12</td>
<td>72</td>
<td>69</td>
</tr>
<tr>
<td>20</td>
<td>56</td>
<td>62</td>
</tr>
<tr>
<td>24</td>
<td>59</td>
<td>62</td>
</tr>
<tr>
<td>26</td>
<td>63</td>
<td>53</td>
</tr>
<tr>
<td>Mean</td>
<td>61.3</td>
<td>59.0</td>
</tr>
<tr>
<td>St. Deviation</td>
<td>6.8</td>
<td>5.7</td>
</tr>
</tbody>
</table>
one or both of the two ratios—current ratio and total debt/total assets—we attempted to generate classifications using rules of thumb which are independent of the composition of our sample based on these two ratios.

A known rule of thumb in traditional fundamental analysis of finance suggests that a current ratio smaller than 2.0 can imply a liquidity problem, which is reasonably confirmed by Beaver's data [1966, p. 82]. Similarly, Beaver's data also showed that failed firms had a total debt/total assets greater than 0.50, while nonfailed firms had a stable ratio of about 0.37. This is equivalent to stating that total debt/equity is 1.00 or higher for the failed firms. Accordingly, two rules of thumb were adopted: (a) classify as defaulted if current ratio is \( \geq 2.0 \); and (b) classify as defaulted if total debt to equity ratio is \( \leq 1.00 \).

The results showed 0.84 and 0.72 correct predictions for the first and the second rule, respectively. Both results are superior to the 0.625 obtained by loan officers in this study, especially given that these rules are independent of the actual composition of default and no-default in our sample, and that these rules use a smaller set of cues compared to any of the models discussed above. Of course, the sample sensitivity of these results is not known.

7. Summary and Limitations

In his synthesis of the 1976 Conference on Human Information Processing, Einhorn suggested that improvements in research in this area could be accomplished by introducing information-seeking behavior which would tend to induce decision makers to generate competing hypotheses about the benefits of each information cue. The main objective of our study was to introduce the concept of sharing information choice with the subjects—a step in the direction of implementing Einhorn's suggestion.

The results can be summarized as follows: (1) On the average, loan officers performed worse than the mechanical models used to generate various benchmarks, including the environmental models. (2) If one assumes that predicting the environmental event is the principal task, the different indices generated by comparing different strategies of information choice and processing indicate that information choice by subjects is the major contributing factor to the relatively low performance of human judges. (3) Unlike the finding about information choice, the utilization of information by human subjects was not materially different from the model, even though the model still outperformed humans. (4) Replication of the study indicated that the knowledge of the composition of the sample (changes in priors) did not affect the quality of judgments made in this study—an unexpected finding. (5) The evidence suggests that judgments of reasonable quality could be made on the basis of the information contained in a few data items. (6) The main result here is consistent with the conclusion drawn by Dawes and Corrigan [1974, p.
that "The whole trick is to decide what variables to look at and then to know how to add." That is, information choice, not weighting, is what matters.

Some limitations of this study must be mentioned. Neither the sample of the decision makers nor the sample of firms were randomly selected. The sample of firms was relatively small since we chose not to ask the participants to make more than thirty-two judgments each in each round of information choice, use, and judgment. Furthermore, we do not know the method by which the subjects made their information choice and whether the choice was made by elimination or by representation (Slovic et al. [1977, pp. 4-7]). No feedback was provided to subjects during the experiment, but the impact of this limitation on the findings is unknown. Finally, it was assumed that the model and the subjects are operating under the same type of loss function. Naturally, the loss function of human subjects may not be symmetric with respect to Type I and Type II prediction errors. These issues could provide a useful basis for further studies on the subject.

REFERENCES


