INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Goldsberry, Betty Sanders

THE EFFECTS OF FEEDBACK AND PREDICTABILITY ON JUDGMENT

*Rice University*  
Ph.D. 1984

University Microfilms International  
300 N. Zeeb Road, Ann Arbor, MI 48106
RICE UNIVERSITY

THE EFFECTS OF FEEDBACK AND PREDICTABILITY ON JUDGMENT

by

BETTY S. GOLDSBERRY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE:

William C. Howell
Professor of Psychology, Chairman

David M. Lane
Associate Professor of Psychology

Robert J. Harvey
Assistant Professor of Psychology

Richard J. Stoll
Associate Professor of Political Science

Houston, Texas

May, 1984
Abstract

THE EFFECTS OF FEEDBACK AND PREDICTABILITY ON JUDGMENT

Betty S. Goldsberry

Previous research has found that when subjects are given cognitive feedback, they reach higher levels of achievement than when they are given outcome feedback. It was hypothesized that this finding was due in part to the predictability of the task environment since outcome feedback is at a distinct disadvantage as a sole means of conveying such information. A study was conducted to compare response and outcome feedback under three conditions which varied in terms of the predictability between actual and optimum criteria. The design included a control group receiving no feedback at all, two response groups differing in precision of feedback information, and two outcome feedback groups differing on a quantity dimension.

Task predictability conditions averaged across five learning blocks were high ($r = .94$), moderate ($r = .87$) and low ($r = .71$). The study also attempted to clarify the definition of feedback and to equate the availability of task information in the various feedback conditions that were compared.

The results, however, did not support the above hypothesis. The utility of outcome feedback was inferior to
that of response feedback under all three predictability conditions tested. In fact, an interaction revealed that the effect of increased predictability raised rather than lowered the disparity between outcome and response feedback performance. Generally, a decline in task predictability accompanied a decline in performance measured in terms of achievement, hit-rate, knowledge, and control. The results also revealed that a control group that received no feedback at all performed as well as or better than those that received feedback when the availability of task information was equated. Moreover, eliminating the memory requirement inherent in the use of outcome feedback only worsened performance. Similarly, adding precision to the response feedback condition beyond the level of mere directional error information did not improve performance.

The principal conclusions to be drawn from these findings are: (a) increasing predictability improves judgment performance, (b) providing outcome feedback is more detrimental to performance than providing response or no feedback when a valid task structure is available, and (c) increasing predictability does not reduce the disparity between the effectiveness of outcome and response feedback.
Dedication

To my parents, Benjamin and Myrtle Sanders
Acknowledgments

This research was supported by the Engineering Psychology Programs, Office of Naval Research, ONR Contract N00014-82-C-0001, Work Unit NRL97-074. The author would like to thank the members of the thesis committee for their constructive comments and suggestions. Special thanks are due to Dr. William C. Howell and Dr. David M. Lane for their assistance during this research endeavor. Special thanks are also due to her sister, Phymeon S. Jackson, whose assistance in the preparation of this manuscript was invaluable to its completion. Finally, the author is thankful to her husband, Ronald, and children, Ryan and Renee, for their patience and encouragement during the pursuit of this degree.
List of Tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Optimum Weighting Strategy</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>Optimum Achievement Scores by Blocks</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>Optimum Hit-rate Scores by Blocks</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>Mean Process Measures by Predictability</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>Mean Product Measures by Feedback Type</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>Mean Process Measures by Feedback Type</td>
<td>60</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Brunswik Lens Model.</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Profile Format</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>Historical Outcome Feedback Display</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>Exact Response Feedback Display</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>Comparative Response Feedback Display</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>Actual Performance by Predictability</td>
<td>51</td>
</tr>
<tr>
<td>7</td>
<td>Relative Performance by Predictability</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>Actual Achievement by Blocks</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>Actual Hit-rate by Blocks</td>
<td>65</td>
</tr>
<tr>
<td>10</td>
<td>Knowledge by Blocks</td>
<td>67</td>
</tr>
<tr>
<td>11</td>
<td>Control by Blocks</td>
<td>68</td>
</tr>
<tr>
<td>12</td>
<td>Actual Hit-rate by Predictability</td>
<td>70</td>
</tr>
<tr>
<td>13</td>
<td>Actual Achievement by Predictability</td>
<td>71</td>
</tr>
<tr>
<td>14</td>
<td>Relative Hit-rate by Predictability</td>
<td>73</td>
</tr>
<tr>
<td>15</td>
<td>Relative Achievement by Predictability</td>
<td>74</td>
</tr>
</tbody>
</table>
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Brunswik Lens Model</td>
<td>5</td>
</tr>
<tr>
<td>Multiple Regression Analysis</td>
<td>8</td>
</tr>
<tr>
<td>Literature Review</td>
<td>11</td>
</tr>
<tr>
<td>Feedback Types</td>
<td>12</td>
</tr>
<tr>
<td>Feedback Effects</td>
<td>17</td>
</tr>
<tr>
<td>Rationale</td>
<td>28</td>
</tr>
<tr>
<td>Method</td>
<td>33</td>
</tr>
<tr>
<td>Subjects and Design</td>
<td>33</td>
</tr>
<tr>
<td>Task</td>
<td>33</td>
</tr>
<tr>
<td>Procedure</td>
<td>41</td>
</tr>
<tr>
<td>Measures</td>
<td>42</td>
</tr>
<tr>
<td>Analysis</td>
<td>44</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>48</td>
</tr>
<tr>
<td>Conclusions</td>
<td>75</td>
</tr>
<tr>
<td>Reference Notes</td>
<td>78</td>
</tr>
<tr>
<td>References</td>
<td>79</td>
</tr>
</tbody>
</table>
Introduction

Studies of learning and judgment have frequently focused on predictive behavior. Subjects are asked to tell which one of a number of outcomes will occur, and the accuracy of their responses is taken as an indication of their knowledge of the task. Often they are provided with information about the outcome of prior behavior in the hope of improving their predictions. This scenario emphasizes a particular order among events. First, a response is made; second, information about the accuracy of that response is presented; then, something happens to the judge that results in an improved response the next time around.

A variety of general terms has been used to describe the procedure of providing individuals with information about the accuracy of prior responses: (a) "knowledge of results" (Dees & Grindley, 1951), (b) "feedback" (Smode, 1958), (c) "reinforcement" (Suppes & Frankmann, 1961), and (d) "reward" (Noble & Alcock, 1958). There has also been a number of specific terms created, usually by adding modifiers to one of the general terms. The modifiers can denote particular environmental manipulations (e.g., "intermittent") or presumed organismic processes (e.g., "cognitive"). In any case, this procedure (whether it is
called knowledge of results, feedback, reinforcement, or reward) has been widely studied as a method of improving predictive performance in human judgment.

The term "feedback" has traditionally referred to some signal occurring after or at the time of a response which provides an indication of its correctness, accuracy, or adequacy. It will be used throughout this paper rather than any of the other labels because it is more neutral in terms of implied processes, and it is currently more widely used (perhaps for that reason). In the laboratory, it is identified as the post-response signal supplied by an experimenter indicating something about the incorrectness of a response or the correctness of another response. Whether either of these kinds of information is necessary for learning has not been clearly established; however, results do point to the sufficiency of information about correctness of a response for learning under some circumstances (Holding, 1959).

Bilodeau (see Bourne, 1966) asserted that feedback expedites learning because it is response strengthening, response sustaining, and error eliminating. Two hypotheses have been offered to explain how this is accomplished. One suggests that feedback operates directly to establish and strengthen the associations between stimulus cues and responses. These cues may be external
or internal. The other hypothesis suggests that feedback strengthens associations as an indirect consequence of some mediating stimulus-response connections which are either tied to or a product of feedback.

It has been documented repeatedly that humans have limited judgment capabilities. Several deficiencies such as heuristic processing, inconsistencies, and misperceptions result from this limited capability (Nisbett & Ross, 1980). The problem is further complicated by the fact that humans seem to be unaware of their limitations. This tendency manifests itself in greater certainty about assessments, theories, conclusions, and events than can be justified objectively. The empirical evidence (see review by Einhorn & Hogarth, 1981; Slovic, 1982; Slovic, Fischhoff, & Lichtenstein, 1977) lends strong support to this generalization. In view of the presumed efficacy of feedback for improving task performance, it stands to reason that the presentation of feedback might serve to reduce these deficiencies as well. In fact, a variety of feedback manipulations has been explored for just this purpose, as we shall see in a moment.

Recent trends in the study of feedback applied to judgment have redirected attention from the judge to the environment (Payne, 1980). More specifically, experiments
are now focusing on the characteristics of the task environment that influence feedback efficacy rather than on the human deficiencies that limit it. The results to date suggest that the effects may be quite systematic although not necessarily simple (Adelman, 1981). If one accepts the notion that both feedback type and task characteristics can play a role and further, that the effect is probably interactive (i.e., certain kinds of feedback work well in some situations but not in others), the problem facing the researcher becomes complex indeed. At the present time, very little is known about such interactions. At one time, for example, it was believed that outcome feedback (indicating the correctness of a judgment) was of little value to the judge; only cognitive feedback (indicating the error in the policy that produced the judgment) could lead to improved performance (Hammond & Summers, 1965, 1972). Now, however, it appears that the usefulness of outcome feedback depends on the congruence of the task (Adelman, 1981).

The present research, therefore, was designed to investigate the effects of feedback on judgmental accuracy in conjunction with one important task variable. Feedback was manipulated in accordance with classical distinctions so that results could be compared meaningfully with earlier research on the topic. The task variable was the
level of predictability afforded by the environment. This was accomplished through the use of a common and versatile judgment model, the Brunswik Lens Model, and an equally versatile statistical procedure, multiple regression analysis. A brief description of each follows.

The Brunswik Lens Model

The "lens model" (illustrated in Figure 1) separates characteristics of the environment from characteristics of the judge. The left portion of Figure 1 represents the environment and illustrates the relationship between the cues (features of the stimulus) and the criteria (the correct responses). The right portion of Figure 1 represents the judge and illustrates the relationship between the cues and the judgments. The left portion, therefore, would permit a normative analysis of judgment while the right portion would permit a descriptive analysis of it.
Figure 1: Brunswik's Lens Model
Tucker (1964) suggested that the relationship between the judgments and the criteria could be partitioned into several statistically independent components reflecting: (a) the judge's acquired knowledge of task properties, (b) his cognitive control in applying that knowledge, (c) the degree of predictability in the task environment, and (d) the nonlinearity in the judgments. The equation reads as follows:

\[
R_a = G R_s R_e + C \sqrt{1 - R_s^2} \sqrt{1 - R_e^2}
\]  

(1)

where

\( R_a \) = the relationship (correlation) between the judgments and the criteria;

\( G \) = the correlation between the linear predictions of the judgments and the linear predictions of the criteria;

\( R_s \) = the correlation between the judgments and the linear predictions of the judgments;

\( R_e \) = the correlation between the criteria and the linear predictions of the criteria, a measure which places an upper limit on achievement;

\( C \) = the correlation between the variance in the task system and the variance in the response system, a measure which is an indication of nonlinearity in the
judge's response strategy.

Hammond and Summers (1972) used the lens model to isolate two of these components—the acquisition of knowledge (G) and the application of knowledge (Rg). They defined acquisition as the extent to which the judge's cognitive system is isomorphic with (in the same form as) the task environment. They defined application, or cognitive control, as the extent to which acquired knowledge is utilized consistently in making judgments. The isolation of these two factors made possible an assessment of how feedback type and task characteristics affect performance.

Multiple Regression Analysis

Assessing the effects of feedback type and task characteristics on human judgment can be accomplished by using multiple regression analysis, a statistical procedure designed to explain the contribution of 2 or more independent variables (cues) to a dependent variable, which could be criteria or judgments. This statistical procedure was developed as a result of the lens model (Brunswik, 1952, 1955) and both the normative and descriptive relationships depicted in Figure 1 can be determined empirically by this procedure.
Multiple regression analysis is used to model the way in which information about the cues is used or should be used to produce a judgment. It accomplishes this by generating a linear regression equation, which depicts how to best weight each cue dimension, from an intercorrelation matrix of cue values and judgments. If the cues are regressed on the correct judgments (criteria), the linear model illustrates an optimum weighting strategy (a normative model of judgment); and if the cues are regressed on the observed judgments, the model illustrates a response weighting strategy (a descriptive model of judgment). If the weight of a cue dimension is 1.0, the judge has relied completely on that dimension in making his judgments but if the weight is zero, the judge has ignored the dimension. This approach to determining weighting strategy offers a suitable index for predicting future judgments. From a practical standpoint, it is important to note, however, that cognitive processing cannot be deduced from these weights (Hobson & Gibson, 1983; Kerkar, 1983).

The present research used the lens model and multiple regression analysis to create a theoretical framework for the investigation of the effects of feedback type and task characteristics on human judgment. It hypothesized that these variables have an interactive
effect on judgmental accuracy. The following section reviews, therefore, the relevant feedback literature in light of this hypothesis.
Literature Review

Most studies examining the effects of feedback on judgment performance have used task environments in which the cues and criteria were continuous or scaled variables rather than discrete ones. These tasks are referred to as metric inference tasks (Bjorkman, 1967). Non-metric inference tasks are considerably different from metric ones because their measures (i.e., colors, letters, and geometric shapes) cannot be ordered along a continuum. Metric inference tasks use measures (i.e., size, scores, and correlations) that require the judge to make discriminations within a dimension rather than across dimensions.

Research has generally revealed that effects of feedback differ to some extent for metric and non-metric tasks. This may be due to the fact that these tasks require different cognitive processing mechanisms. A metric task may require a more precise mechanism than a non-metric one because finer discrimination is required. For example, the selection of a specific orange from a basket containing a variety of fruit could be based on color, size, smell, or shape (all different dimensions of fruit), but the selection of that same orange from a basket of oranges would have to be based on
characteristics generic to some extent to all oranges. The study of metric tasks, therefore, could potentially reveal more about underlying policies and cognitive processing capabilities than the study of non-metric tasks because the dimensions used in the discrimination would have been isolated. For this reason, the scope of the literature review and research that follows has been limited to metric tasks.

Feedback Types

Feedback types are often described in terms of the functions they perform for the judge or recipient. These functions have been described as either having directional or motivational characteristics (Locke et al., 1968). In general, directional characteristics serve the function of informing the recipient about his behaviors while motivational characteristics serve the function of informing the recipient about the outcomes or rewards of his behaviors (Arnett, 1966).

Although feedback performs several functions, it is doubtful that effects of these functions can be completely separated from one another. It is, therefore, efficacious to view feedback in a broader context, such as a continuum of functions ranging from directional to motivational. At the directional end of the continuum,
feedback would provide precise information about how to perform the task; at the motivational end, feedback would provide information about the results of that performance.

Feedback type has also been described in terms of dimensions that may dictate level of cognitive processing (i.e., timing, sign, quantity, and precision). Timing refers to the properties of the interval between the judge's response and the issuance of feedback (i.e., duration or variability). Sign refers to the nature of the relationship between the feedback and the response (i.e., positive or negative). Quantity refers to the amount of information provided at one issuance of feedback. Precision refers to the level of accuracy of the information provided. Other dimensions have also been used to differentiate between feedback types, but these four appear to be the most common.

Research has shown that learning a judgment task is affected by varying feedback along these types of dimensions, and interactions between specific dimensions have been observed. For example, even though positive information is generally perceived and recalled more accurately than negative information (Ilgen, 1971), there is some indication that timing may affect them in different ways. Surber and Anderson (1975) published
results supporting this view. They found that the learning of correct responses on a multiple choice test was better when negative feedback was delayed and when positive feedback was immediate.

Feedback, therefore, has been functionally defined and studied along several different dimensions. One way to integrate these dimensions is to think of feedback as a special case of the communication process in which a source (environment) conveys to a recipient (judge) a message (feedback) about his behavior (performance). The judge's perceptions of and responses to the feedback depends on his personal characteristics as well as the environmental properties. By focusing on the environment, the judge, or the performance, feedback can be categorized as one of three classical types—task feedback, response feedback, or outcome feedback. If it provides information about the structure of the environment, it is task feedback; if it provides information about the response strategy of the judge, it is response feedback; and if it provides information about the results of his judgments, it is outcome feedback.

Now, returning to the previous distinctions, it is apparent that task, response, and outcome feedback can vary along dimensions of timing, sign, quantity, and
precision and that they can be oriented chiefly toward the directional or the motivational functions. The precision of feedback might be considered exact, for example, when it is presented in terms of regression weights, and comparative when only the directional error of the regression weights is given. The quantity of feedback would be considered small when information on only the immediately preceding response is provided, and it would be large when the history of previous responses is provided as well.

Much of the early research in this area combined task and response information under the general category of cognitive feedback because it was believed that both of these types of information were necessary. In this context, task information indicated how the judge should have processed the cues, and response information indicated how he actually did process them. By contrast, outcome feedback indicated nothing at all about processing other than what could have been deduced over time from collective observations. Outcome feedback, therefore, could be looked upon as providing indirect guidance to cognitive processing while cognitive feedback could be looked upon as providing direct guidance to it.

The classical distinction between cognitive and outcome feedback can be related to the functional
distinctions made earlier. That is, feedback is \textit{directional} when it gives the judge information about the task and \textit{motivational} when it gives him information about results. Hence, the principal function served by cognitive feedback is that of providing direction, whereas the principal function of outcome feedback is that of providing motivation.

Two studies (Newton, 1965; Nystedt & Magnusson, 1973) have questioned the appropriateness of referring to information about task characteristics as feedback. Since task information is a description of the task environment and is not directly affected by the responses, they have argued that it is informative but not feedback. Because feedback is defined in the dictionary as "the return of part of the output of a system to the input for purposes of modification and control of future output" (Funk & Wagnalls, 1963), this view has considerable merit. This definition, therefore, would only qualify task information as feedback when it has been altered in some way by previous responses. Knowledge of this alteration could then possibly result in a change in the judge's response strategy and aid future performance.

As this review has shown, there are many overlapping distinctions among the major types of feedback. These distinctions can be used to place cognitive and outcome
feedback on the same continuum, a theoretical perspective supported by Hammond's Cognitive Continuum Theory (1980). His theory proposes that a judgment task evokes a mode of cognitive processing that varies from intuition to analysis depending on characteristics of the judgment task. These characteristics are determined by its structure, content, and presentation. Therefore, since feedback type determines to large extent task content, it may have a systematic effect on cognitive processing. The rest of this literature review will be devoted to summarizing the research on cognitive and outcome feedback and comparing their relative effects on cognition and performance.

**Feedback Effects**

**Cognitive feedback.** As noted earlier, researchers have not always made a clear distinction between task and response feedback. Often, both were manipulated in the same experiment under the general category of cognitive feedback. For example, a study by Hammond (1971), which used computer graphics as an aid in implementing the lens model, provided the subject with graphical and numerical information about the statistical properties of the task and his responses. Task information depicted the relationship between the cues and the criteria while
response feedback depicted the relationship between the cues and the judgments. The results showed that this feedback enhanced the judge's cognitive control.

In another study, a man-computer system called cognograph (Hammond & Summers, 1972) was designed to improve judgments based on information provided prior to the task. Subjects' judgments were fed directly into a computer which generated a graph of optimum and utilized weighting strategies. This form of feedback was also found to be useful in improving the accuracy of subsequent judgments.

Cognitive feedback has taken several statistical forms. Newton (1965) used a variety of measures as feedback, including regression weights (\( R \)) on both criteria and judgments. Regression weights on the criteria are basically task information while regression weights on the judgments are response information. Newton also used correlations between criteria and responses (\( r \)) and verbal rules as cognitive feedback. Adelman (1975) and Brady and Rappoport (1973) used \( R \); Flack and Summers (1971) used \( r \); and Balke et al. (1974) used a composite measure, \( Br/R \). They all reported improved judgments after these types of cognitive feedback. Darlington (1968), however, showed that \( R \) and \( r \) could provide different information to the
judge. He stated further that the simultaneous use of these measures could introduce ambiguity into the judgment task, particularly when cues are correlated.

The few studies that have investigated the effects of task information on performance (Hammond, 1971; Hursch et al., 1964) gave task feedback to one group of judges and cognitive feedback (a combination of task and response information) to another group. Positive effects were found for both groups; however, in these and other similar studies, cognitive feedback was found to provide no greater improvement in performance than task feedback alone (Gillis, Gritz, & Stewart, 1975; Nystedt & Magnusson, 1973). Taken together, these results strongly suggest that the knowledge of the task system, not the response system, may be the important ingredient in cognitive feedback.

Steinmann (1976) used a similar approach to distinguish between the effects of task and response information. His primary goal, however, was to discover why cognitive feedback was no more helpful than task information alone. He proposed that the use of very simple linear relationships in the judgment tasks in which these results were found was responsible. To test his theory, he manipulated task complexity by varying the type of relationship between the cues and criterion. The
relationships varied from a simple linear one to a very complex nonlinear (U-shaped) one. He expected cognitive feedback to be more effective than task information for the more complex relationships but not for the simple ones. However, no theoretically important differences were found between the two. In fact, performance was extremely high for both feedback types and all relationships tested.

The possibility that task information alone is sufficient for learning a judgment task contradicts the original premise on which the use of cognitive materials as feedback was based. In an early report, Hammond and Summers (1972) asserted that in order for feedback to contribute to cognitive control, it had to consist of enough cognitive information for judges to perceive that their judgments were in error as well as why they were in error. Hence, their investigations concluded that cognitively oriented feedback had to include two components: (a) the properties of the cognitive system and (b) the properties of the task system. They believed that a comparison of these two systems leads to learning a judgment task. Not all the research findings, however, support the necessity of this comparison, a point developed further in the next section.

Outcome feedback. With outcome feedback, the judge has
access to information about whether his judgments have led to overall success or failure but no information about why. Research has shown this type of information to be effective in learning academic concepts (Gagne, 1962) and psychomotor skills (Bandura, 1971).

Psychomotor skills seem to rely heavily on an "automatic" type of cognitive processing— one that operates with minimal awareness and attentive effort— while judgment tasks rely on a controlled, attention-demanding, conscious type of cognitive processing (Shiffrin & Schneider, 1977). Since it has been shown that cognitive control declines in the presence of outcome feedback (Hammond & Summers, 1972), one might expect outcome feedback to be more effective in a task requiring less conscious cognitive control.

Evidence from several sources has suggested that judges have difficulty maintaining cognitive control after receiving outcome feedback (Dudycha & Naylor, 1966; Goldberg, 1970, 1971; Hammond & Summers, 1972). This view received more recent support in a study that examined control and matching indices in both outcome and no-outcome feedback conditions (Schmitt, 1978). Outcome feedback had negative effects on control and matching proficiency when both statistical and subjective weights were evaluated.
Viewed in motivational, rather than informational, terms, the functional effects of outcome feedback have been reported to exert a positive influence on learning and performance. In one study, for example, Ilgen (1979) reported that subjects informed of how well they were performing were more interested in learning the task than those who were not so informed. Therefore, supplying some indication of accomplishment in the form of outcome feedback can help to maintain interest, particularly if the task is tedious. The link between motivation and performance, however, has not been clearly delineated.

The utility of outcome feedback observed in concept (Gagne, 1962) and skill (Bandura, 1971) learning, however, has not been duplicated in research on human judgment. In one study (Lichtenstein & Fischhoff, 1980), subjects were trained to make probability assessments and then given outcome feedback to assist them in improving their performance. Most improvements in performance occurred after some training but prior to receiving outcome feedback. Howell and Emanuel (1968) also found that some subjects made their greatest performance improvements prior to receiving outcome feedback. These results raise questions concerning the value of outcome feedback over practice alone, particularly for individuals who are "well-calibrated."
In two studies (Norman, 1974a, 1974b), subjects were given outcome feedback on the accuracy of their judgments in a psychomotor averaging task. This feedback led to changes in both the type of integration rule used and the weighting of the cues presented. The results showed that outcome feedback affected learning, but they did not clarify the nature of the effect. A third study (Norman, 1976) revealed that the observed changes in performance were not necessarily a positive function of the feedback presented. The effect of relevant information increased cognitive activity while the effect of irrelevant information decreased it. Neither of the effects, however, could be linked to specific increases in performance. Results simply showed that the presentation of outcome feedback yielded a systematic change in the weighting of relevant and irrelevant information.

Reasons for the adverse effects of outcome feedback were sought by Fischhoff (1975). He found that outcome feedback led to overestimations of prior knowledge and to changes in the perceived relevance of cues. Simply stated, subjects' perceptions of the data were distorted after receiving outcome feedback: it led them to believe that they knew more than they actually did. In addition, results showed that original predictions of events and later recall of those predictions were poorly correlated
because subjects underestimated low probability events. Fischhoff (1975) called this reaction to outcome feedback "creeping determinism" and described it as a tendency to perceive reported outcomes as having been relatively inevitable. He concluded that creeping determinism hindered the subject's ability to learn from outcome feedback because it caused inappropriate response strategies to be positively reinforced.

Einhorn (1980a) supported Fischhoff's view by suggesting that outcome feedback may be irrelevant to future performance since it tells the subject about the success or failure of his response strategy without telling him why. This issue was discussed within his concept of outcome-irrelevant-learning-structures (OILS). This term suggests that positive outcome feedback can be irrelevant for correcting poor response strategies when the task structure is unknown. The concept of OILS implies that positive outcomes can reinforce the use of inappropriate strategies and keep the subject unaware that his strategy can be improved.

**Cognitive versus outcome feedback.** The relative effectiveness of outcome and cognitive feedback has been investigated extensively over the last several decades (Deane et al., 1972; Hammond & Summers, 1965, 1972; Summers & Hammond, 1966). Cognitive feedback has
generally been found to improve both the acquisition and
and the application of knowledge about task structure,
while outcome feedback has been generally found to hinder
it. Two studies (Lindell, 1965; Todd & Hammond, 1965)
even found that cognitive feedback led to higher
performance than outcome feedback in judgment tasks that
varied the disparity between the relative importance of
the cues.

Adelman (1981) hypothesized that the superiority of
cognitive feedback over outcome feedback reported in the
literature was primarily due to the "neutrality" of
situations present in virtually all of these studies.
Neutral task content was defined as a condition in which
the subject receives no useful information about the
actual structure of the task. Adelman observed that
judgment research had limited task content level to the
point where no genuine understanding of the substantive
properties of the task existed prior to feedback. This
observation led him to question whether cognitive
feedback would be superior to outcome feedback when task
content was non-neutral. Non-neutral task content
suggests that task information may be positive, providing
a valid assessment of the cue-criterion relationship in
the environment, or negative, providing an invalid one.
Findings revealed that non-neutral task content yielded
similar performance for cognitive and outcome feedback but that neutral task content yielded superior performance for response feedback.

The differential effects of cognitive and outcome feedback is thus an unresolved issue in human judgment. In fact, Hoffman et al. (1981) recently developed a multidimensional approach to its study in which subjects had active control over the presentation of a general class of multidimensional stimulus patterns which were functionally related to a criterion. Their findings supported the utility of cognitive feedback but not outcome feedback. In addition, their results illustrated that a new type of continuous feedback called partial differential feedback (PDF) enabled subjects to learn more complex functional relationships than had been thought possible.

Partial differential feedback is cognitive feedback presented in a judgment task environment characterized by a subject-controlled stimulus and a tonal-frequency criterion. The actual feedback presented is the partial derivative of the functional relationship between the stimulus and the criterion. Its function, therefore, is to communicate the rate of change in the criterion to changes in the stimulus.

In summary the issue regarding the relative efficacy
of outcome and cognitive feedback in judgment tasks is far from resolved and, therefore, continues to stimulate research in the area of human judgment.
Rationale

Useful task property information defines the judgment task environment. It identifies the relevant features of the stimulus (cues) and the criterion. For example, if some index of fine motor coordination (a cue) is relevant to some index of success as a surgeon (criterion), the relationship between the two indices would be useful information to have when selecting candidates for this medical speciality. Of course, the index of fine motor coordination would probably be but one of several cues necessary to make an accurate judgment about success as a surgeon.

As noted in the last section, recent research has focused on the possibility that task property information may have a significant effect on the relative utility of cognitive and outcome feedback (Adelman, 1981). Although Adelman's study dealt with only one task property, congruence, research on other task properties has led to a similar conclusion (Hammond & Summers, 1965; Schmitt et al., 1977; Summers & Hammond, 1966). A reasonable approach to investigating the generality of this effect would be to identify meaningful properties and then examine the systematic manipulation of them on feedback
efficacy. This, of course, is what Adelman did with the congruence dimension. Task congruence is a measure which indicates for a group of judges the correspondence between implied and actual properties of the task environment. It was operationalized in his research as the average correlation obtained, without feedback, between predictions derived from the regression models of the responses and the criteria. This measure is also the average knowledge parameter (\( G \)) in the lens model equation when obtained without feedback.

Another dimension, task predictability, has also been suggested as a possible factor in feedback efficacy (Payne, 1980). It is defined as the reliability of the optimum relationship (weighting strategy) and is operationalized as the correlation between the criterion and the linear prediction of the criterion. Since Adelman (1981) found an interaction between task congruence and feedback type, one might expect to find a similar relationship between task predictability and feedback type.

The present study, therefore, was designed primarily to investigate the influence of the task property, predictability, on feedback efficacy. Task predictability was set at 3 discrete levels—high (\( r = .94 \)), moderate (\( r = .87 \)), and low (\( r = .71 \)). It was lowered by
progressively adding random error or "noise" to the cue-criterion relationship present in the task environment. Based on the results of Adelman's research, a feedback type by predictability interaction was expected such that the difference between the effectiveness of outcome and cognitive feedback would change as an inverse function of predictability. More specifically, performance based upon response feedback was expected to exceed that based upon outcome feedback, but the difference between the two was expected to decrease significantly with an increase in task predictability.

Two types of feedback were investigated in this study. One type involved cue-utilization information (response feedback), and the other involved criterion information (outcome feedback). Response feedback was presented at two levels of precision: exact regression weights or comparative information about the direction of weighting errors. The purpose of this manipulation was to determine what aspects of the feedback information subjects actually use in establishing their weighting policy. A significant effect would suggest that they are able to incorporate the metric information into their judgments, while no effect would suggest that they respond mostly to the comparative aspects of it.

Outcome feedback was also presented at two levels
but here the manipulation was the number of prior judgments about which results were displayed: the one profile just completed versus the past 20 profiles. This manipulation was designed to investigate the role of memory in the adverse results found for outcome feedback. Significantly higher performance when the quantity of information is increased would suggest that subjects have performed poorly after outcome feedback because they could not recall enough about previous outcomes to benefit from this information.

All subjects were instructed to base their judgments on an optimum weighting strategy which informed them of the cue-criterion relationships in the task environment. They were also told how predictable this strategy would be in selecting the criterion if it was applied consistently. By providing an "ecologically" valid relationship, their proficiency in using this information could be evaluated. It also created a realistic environment in which to make judgments. Consider the possibility of asking someone to diagnose illnesses with no knowledge of medicine, or to hire new employees with no knowledge of the job. When subjects are not provided with information about the cue-criterion relationships in the task environment, the primary focus of the investigation is on how rapidly they can learn; when they are provided with this information,
the focus is on how accurately they can perform. This research, therefore, investigated whether judgment performance varies with different kinds and levels of feedback as the "true" task structure is progressively obscured by "noise."
Method

**Subjects and Design**

Seventy undergraduate psychology students participated in the research project as judges (subjects) in exchange for extra course credit or $12.00 in cash. Each subject was randomly assigned to one of five treatment groups defined on the basis of feedback type. They made judgments on three different hypothetical jobs. The distinction between these five feedback groups and three jobs will be explained in detail in the next section.

To control for possible order effects, presentation of the three jobs was counterbalanced; to minimize fatigue effects, each job was presented at a different session. Each session was divided into a warm up period and five practice blocks. The experimental design, therefore, was a mixed model 5 (feedback type) by 3 (task predictability) by 5 (practice block) factorial with 14 subjects per group.

**Task**

The judgment task was chosen on the basis of its common usage in human judgment research and the likelihood that it would be meaningful for a wide variety of
potential subjects. It consisted of evaluating the suitability of 320 hypothetical job applicants using as predictors scores on nine-point skill ratings. Based on the optimum weighting strategy (Table 1), subjects were required to integrate profile information for each applicant into a single suitability rating on a scale of 1 (least suitable) to 9 (most suitable).

**TABLE 1**

Optimum Weighting Strategy

<table>
<thead>
<tr>
<th>Skill Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression Weight</td>
<td>.50</td>
<td>.30</td>
<td>.20</td>
</tr>
</tbody>
</table>

The optimum weighting strategy is a normative model of the task environment (left portion of the lens model) and was used to generate the criteria and the criterion predictions for this judgment task. Random error was then added to the criteria to establish the three levels of predictability (one for each job): (a) high, in which \( r = .94 \), (b) moderate, in which \( r = .87 \), and (c) low, in which \( r = .71 \). Therefore, three different sets of criteria
were generated and only one set of predictions. In the low predictability condition, 50 percent of the variance was due to random error; in the moderate and high predictability conditions, the corresponding error variances were 24 and 12 percent, respectively. These three sets criteria were used as a basis for generating the feedback in all four of the groups in which it was given.

Each applicant profile (illustrated in Figure 2) contained three skill ratings and a set of irrelevant biographical information. It was presented via a Visual 200 terminal controlled by an Advanced Systems/9000 computer. The skill ratings were generated orthogonally from a normal distribution of numbers ranging from 1 to 9 with a mean of 5 and a variance of 2. Biographical data were randomly selected from the Houston, Texas telephone directory.

The five treatment groups defined on the basis of feedback type were: (a) no feedback (control), (b) non-historical outcome feedback (criterion for the immediately prior judgment), (c) historical outcome feedback (a record of the past 20 judgments), (d) exact response feedback (utilization regression weights for the last 20 profiles), and (e) comparative response feedback (direction of the weighting error for the last 20
Name: Mary Frances Smith
522 Pontiac Avenue
Houston, Texas 77024

Telephone No. 567-3443

Rating Skill No. Rating
1 7
2 2
3 5

Your response is _____.

Figure 2: Profile format.
profiles).

Outcome feedback was generated by applying the optimum weighting strategy to the cue ratings. Subjects receiving non-historical outcome feedback were presented the message: "The correct response is #." after each judgment. The symbol "#" denotes the criterion or what the correct response should have been. Subjects receiving historical outcome feedback were also presented this message after each response plus a history of judgments at the end of a 20-profile unit. This history (illustrated in Figure 3) linked each judgment with the corresponding cue ratings and criterion.

Response feedback was created by regressing ratings onto the judgments after each set of 20 profiles (three per learning block). The generated regression weights were then used to create the response feedback displays. These regression weights denoted the subject's utilization of the skill dimensions and were presented along with the optimum regression weights in the exact feedback display (illustrated in Figure 4). In the comparative response feedback condition, the subjects' regression weights were compared with the the optimum weights and only the direction of the weighting error was presented (illustrated in Figure 5). The "OK" message was used when a subject's weight for a particular skill dimension fell
<table>
<thead>
<tr>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Your Response</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 3: Historical outcome feedback display.
<table>
<thead>
<tr>
<th>Rating</th>
<th>Skill</th>
<th>Optimum Weighting</th>
<th>Your Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>.50</td>
<td>.67</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.30</td>
<td>.27</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.20</td>
<td>.06</td>
</tr>
</tbody>
</table>

Figure 4: Exact response feedback display.
<table>
<thead>
<tr>
<th>Rating Skill</th>
<th>Optimum Weighting</th>
<th>Your Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.50</td>
<td>Too high</td>
</tr>
<tr>
<td>2</td>
<td>.30</td>
<td>OK</td>
</tr>
<tr>
<td>3</td>
<td>.20</td>
<td>Too Low</td>
</tr>
</tbody>
</table>

**Figure 5:** Comparative response feedback display.
within a ± .05 interval of the optimum for that dimension.

**Procedure**

All subjects participated in three sessions, each approximately 60 minutes in length and scheduled one week apart. At the beginning of each session, written instructions were given describing the job, the assessment procedure, the feedback type, and the optimum weighting strategy. To insure full understanding of this information, instructions were augmented by a graphical illustration of how each skill dimension correlated with on-the-job performance. The subjects were also taught how to use the Visual 200 terminal to enter their judgments.

Subjects were told that each session would be devoted to making suitability ratings on applicants for three different jobs. The normality and dependence features of the profiles were also explained so the subjects would not be misled into searching for nonexistent profile structures. The actual job titles and skill dimensions, however, were not identified so that subjects would not be influenced by prior knowledge or familiarity with the jobs.

All subjects were paced through 20 warm up profiles followed by 300 experimental ones with the aid of a tape recording of "beeps" presented at 10-second intervals.
Between each set of 20 profiles (a unit) there was a 60-second pause during which the control and non-historical groups rested and the other three groups received their end-of-unit feedback.

**Measures**

The scaled judgments served as the basis for two sets of derived measures: two product measures (hit-rate and achievement) and two process measures (knowledge and cognitive control). The product measures indicated how closely judgment performance approximated an optimum, while the process measures examined the inferred cognitive elements underlying that performance. Computation of product measures required definition of an optimum model relating cues (skill dimensions) to a criterion. In other words, an "ecologically" valid relationship must exist before one can study proficiency.

The model adopted in this research was the standard linear regression approach commonly used in human judgment research and explained in the Introduction section. The optimum weights assigned to the various skill dimensions have already been explained and illustrated in Table 1. Hit-rate, was simply the proportion of a subject's judgments that matched the output of the "true" or ecologically valid weighting model. Achievement was the
correlation between the subject's judgments and the optimum model's "judgments." Knowledge, or the subject's understanding of the optimum weighting strategy, was indexed by the correlation between the criteria produced by the optimum weighting strategy and judgments produced by a model of the subject's weighting strategy. The latter, of course, was derived from the subject's actual judgments through the use of multiple regression analysis to "capture" his policy. Control was indexed by the correlation between judgments predicted on the basis of the subject's captured policy and those actually produced by him. The mathematical relationship between achievement and the process measures (knowledge and control) was discussed in the Introduction. For the conditions present in this experiment, this relationship, which is a mathematical statement of the lens model, is simplified to the following equation:

\[ R_a = G \cdot R_s \cdot R_e \]  

(2)

because specification of an optimum weighting strategy makes the criterion as predictable as \( R_e \) and the optimum strategy is linear (eliminating the need for the right-most term of the equation 1). The simplified equation renders the distinction among measures used in
the present research apparent. The $R_a$ term represents achievement which can be partitioned into knowledge ($G$), control ($R_s$), and predictability ($R_e$). Judgments are, therefore, accurate ($R_a$) to the extent that they correspond with the actual suitability of the applicants as reflected by the substantive properties of the task. It follows, therefore, that a subject can be accurate in his judgments to the extent that he has a predictable task structure ($R_e$), he understands the structure ($G$), and he is capable of using that understanding consistently ($R_s$).

**Analysis**

Product and process measures were determined in keeping with the definitions just presented. Since task predictability is a component of performance as well as an independent variable in this research, actual and relative measures of performance were calculated and analyzed. The relative performance measures were calculated by dividing each observed block score by the optimum block score (obtained by applying the optimum weighting strategy to the cue values). Naturally, the optimum block score declined as predictability was reduced. It also varied somewhat across learning blocks since predictability was not counterbalanced over blocks in this study. Tables 2
and 3 show the optimum block scores used to calculate the relative achievement and hit-rate measures, respectively.

Separate analysis of variance procedures were performed on actual and relative measures in order to assess the significance of main effects and interactions among the independent variables. Dunnett tests, which compare treatment means to a control group, were also used to compare the effects of each feedback type with no feedback at all. In addition, Newman-Keuls tests of paired comparisons were carried out, when appropriate, to isolate the pattern of significant effects. These tests were performed in a manner described by Winer (1962).
Table 2
Optimum Achievement Scores
Total Job Task Predictability (r)

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.83</td>
<td>.87</td>
<td>.72</td>
<td>.81</td>
</tr>
<tr>
<td>2</td>
<td>.95</td>
<td>.91</td>
<td>.80</td>
<td>.87</td>
</tr>
<tr>
<td>3</td>
<td>.97</td>
<td>.87</td>
<td>.61</td>
<td>.82</td>
</tr>
<tr>
<td>4</td>
<td>.98</td>
<td>.92</td>
<td>.78</td>
<td>.90</td>
</tr>
<tr>
<td>5</td>
<td>.96</td>
<td>.86</td>
<td>.64</td>
<td>.82</td>
</tr>
<tr>
<td>Mean</td>
<td>.94</td>
<td>.87</td>
<td>.71</td>
<td>.84</td>
</tr>
</tbody>
</table>
Table 3

Optimum Hit-rate Scores
Task Predictability Levels (% correct)

<table>
<thead>
<tr>
<th>Blocks</th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>63</td>
<td>32</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>62</td>
<td>40</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>50</td>
<td>32</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>63</td>
<td>45</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>65</td>
<td>50</td>
<td>72</td>
</tr>
<tr>
<td>Mean</td>
<td>100</td>
<td>60</td>
<td>40</td>
<td>67</td>
</tr>
</tbody>
</table>
Results and Discussion

For purposes of clarity in exposition, the principal findings are organized around four major questions addressed by this research: (a) Does task predictability influence performance? (b) Does feedback type influence performance? (c) Does practice influence performance? and (d) Does feedback efficacy differ as a function of task predictability? The data bearing on each of these questions will be preceded by a brief discussion of the expected results and their relevance to human judgment. Findings will be based on analyses of variance performed on the actual hit-rate \( H_a \), relative hit-rate \( H_r \), actual achievement \( R_a \), relative achievement \( R_r \), knowledge \( G \), and control \( P_s \) measures of performance which were discussed in detail earlier.

Feedback types will be represented on the figures in this section by the following abbreviations: (a) T, for the control group given only task information, (b) N, for the non-historical outcome group given the correct response after each judgment, (c) H, for the historical outcome group given the preceding correct response and a history after 20 profiles, (d) C, for the comparative response group given the direction of weighting error after 20 judgments, and (e) E, for the exact response group given
regression weights of cue utilizations after 20 judgments.

1. Does task predictability influence performance?

Task predictability can have two different types of influence on performance. First, as an independent variable and means of manipulating a substantive property of the judgment task, predictability can alter the performance of the optimum weighting strategy. A reduction in predictability, therefore, should produce a performance decrement even if the subject makes accurate and consistent usage of the available task information. Second, predictability can have an effect on the way in which subjects process the information presented to them. This, of course, is the more interesting influence from a psychological standpoint. By contrast, the first influence is important in that it serves as an indication of the subject's sensitivity to the manipulation—in essence a method check.

As noted in the Method section, two different kinds of performance measures (actual and relative) were used to explore the effects of predictability on judgment. The influence of predictability as a manipulator of the task content was investigated by analyzing the actual hit-rate and achievement measures. A significant effect for these measures would suggest that subjects were trying to use the
optimum weighting strategy to make their judgments or at least that they were sensitive to the manipulation. The influence of task predictability as a cognitive component of performance was investigated by analyzing the relative performance measures since they reflect performance after manipulation effects have been removed. Significant effects for both actual and relative measures were anticipated.

In general, the predictability manipulation had the desired effect on performance, for as predictability increased, so did performance. Although this finding comes as no surprise, it is nonetheless important because it suggests that subjects were sensitive to this task property and were trying to maximize their performance. As illustrated in Figure 6, both product measures of performance yielded significant effects for task predictability, \( F (2, 130) = 672.15, \ p < .01 \) for actual achievement, and \( F (2, 130) = 377.70, \ p < .01 \) for actual hit-rate.

Predictability was also found to have a significant effect on relative performance. The two relative indices did not agree, however, on the nature of this effect. Figure 7 indicates that relative achievement increased with an increased in predictability, \( F (2, 130) = 8.01, \ p < .01 \) and that relative hit-rate decreased with an
increase in predictability, $F(2, 130) = 45.63, p < .01$.

The ambiguity suggested by the relative measures could be an artifact of the way accuracy was defined for hit-rate: only an exact match between the subject's judgment and the value produced by the optimum weighting model was considered a "hit." Consider the optimum hit-rate values shown in Table 3. Under high predictability, the model was 100 percent correct by definition. Since even a well-calibrated subject would have trouble approximating this level, given that a judgment had to be perfect to be considered correct, one would expect relative hit-rate to be rather low. Under low predictability, on the other hand, the model was correct only about 40 percent of the time, a considerably easier standard against which to express the subject's performance. Even random responding would have yielded a relative hit-rate of nearly 25 percent under this condition (compared to about 11 percent under high predictability). Had a more lenient criterion been set for the definition of a "hit," such as a one unit confidence interval around the model's judgments, relative hit-rate might have produced a trend more similar to that of relative achievement.

The process measures, knowledge and control, provide insight into the manner by which cognitive aspects of judgment affect performance. As illustrated below in Table
TABLE 4

Mean Process Measures under
Three Predictability Conditions

<table>
<thead>
<tr>
<th>Predictability Level</th>
<th>Measures (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Knowledge</td>
</tr>
<tr>
<td>Low</td>
<td>.88</td>
</tr>
<tr>
<td>Moderate</td>
<td>.89</td>
</tr>
<tr>
<td>High</td>
<td>.89</td>
</tr>
<tr>
<td>Mean</td>
<td>.89</td>
</tr>
</tbody>
</table>
4, they reveal that when predictability is increased from the low to the moderate level, both the understanding and the application of task structure information increases; but when it is increased further to the high predictability level, no additional improvement in cognitive processing occurs. Although small in absolute terms, the difference between low and moderate predictability yielded a significant effect of task predictability for both knowledge, $F(2, 130) = 4.92, p < .01$, and control, $F(2, 130) = 3.70, p < .03$. Hence an increase in task predictability can significantly improve these two aspects of cognition; but if they are already at a limit dictated perhaps by mental capability or capacity, no further improvement will occur.

Task predictability, therefore, had a significant, though very small, effect on judgmental knowledge and control. As predictability increased, so did these aspects of judgment, although perhaps limited by a "ceiling" associated with the particular properties of the task.

2. Does feedback type affect performance?

The present study was designed to control for the typical confounding of task knowledge and feedback type. In previous studies, only subjects receiving cognitive feedback had access to specific (and important) knowledge
of the formal task structure. By providing such information to all feedback groups via instructions, the present design permitted a fair comparison of response and outcome feedback conditions. The comparison of each feedback type with a control group receiving no feedback (task information only), therefore, isolated its utility in judgment. In light of recent findings, all feedback types tested were expected to provide some benefit, at least under some conditions, with response feedback generally being more beneficial than outcome feedback.

The results revealed that the type of feedback presented had a significant effect on performance for both actual indices, \( F (4, 65) = 7.82, p < .01 \) for achievement and \( F (4, 65) = 4.21, p < .01 \) for hit-rate. However, contrary to expectations, Table 5 illustrates that the two response feedback groups performed at about the same level as the control group, but all three were better than the two outcome feedback groups. This relationship was supported by results of Dunnett tests which revealed that for relative measures, performances for exact and comparative groups were not significantly different from the control \( (p > .05) \). However, performance for historical and non-historical groups were significantly inferior to the control \( (p < .05) \). This finding generally challenges the utility of the four types of feedback tested when subjects
Table 5

Mean Product Measures under Five Feedback Conditions

<table>
<thead>
<tr>
<th>Measures</th>
<th>Actual</th>
<th>Relative (ratios)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$F_a$ (%)</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>Feedback Types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>50</td>
<td>.76</td>
</tr>
<tr>
<td>Outcome feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-historical</td>
<td>43</td>
<td>.71</td>
</tr>
<tr>
<td>Historical</td>
<td>40</td>
<td>.67</td>
</tr>
<tr>
<td>Response feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative</td>
<td>47</td>
<td>.75</td>
</tr>
<tr>
<td>Exact</td>
<td>50</td>
<td>.75</td>
</tr>
<tr>
<td>Mean</td>
<td>47</td>
<td>.73</td>
</tr>
</tbody>
</table>
are informed of the proper weighting strategy prior to their judgments. It does, however, corroborate the common finding that outcome feedback is detrimental to judgment performance.

It is apparent from the relative performance measures that when the effect of predictability as a manipulator of task content is removed, the relationships between feedback types are unchanged. Significant feedback type effects were found from both $R_X, F(4, 65) = 7.51, p < .01$, and $R_X, F(4, 65) = 2.86, p < .03$.

A comparison of the two types of outcome feedback suggests that presenting a history (consisting of cues, judgments, and criteria for the past 20 profiles) at the end of a profile unit is more detrimental to performance than not presenting such a history. In other words, a large quantity of outcome feedback appears to be more detrimental than a small one. This finding was supported by a Newman-Keuls test that revealed a significant difference in actual achievement for both outcome feedback groups ($p < .05$). Both measures were lower when a history was presented, but the difference was significant only for actual achievement.

It will be recalled that outcome feedback was manipulated in order to determine whether deficiencies of memory could explain the decline in performance observed after outcome feedback in earlier studies. If the decline
was due to failure of memory, performance should have been better with a history of outcomes than without one. Therefore, since a history hurt performance in this study, failure of memory would not seem to explain the decline.

The detrimental effect of historical outcome feedback can, however, be linked to cognitive processing. Inspection of the knowledge and control measures on Table 6 reveals that the subjects receiving a history of their judgments exhibited less control over their response strategies than did subjects who did not receive a history. A Newman-Keuls test supported this conclusion, the difference between these two outcome feedback conditions being significant ($p < .01$) for the control index. In fact, control dropped .06 units (from $R_s = .86$ to .80) when a list was added to the design. No difference in knowledge was observed.

Turning to the more constructive feedback types, it appears that feedback with the precision of regression weights leads to no better performance than comparative information based on those regression weights. This is supported by the fact that a Newman-Keuls test yielded no significant difference ($p > .05$) for the two response groups on any of the measures analyzed. Regression weights, therefore, must be simplified to some extent during cognitive processing. This research does not permit speculation on the degree of simplication that takes place
Table 6

Mean Process Measures under Five Feedback Conditions

<table>
<thead>
<tr>
<th>Feedback Type</th>
<th>Measures (r)</th>
<th>Knowledge</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>.90</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>Outcome feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-historical</td>
<td>.87</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Historical</td>
<td>.87</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Response feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative</td>
<td>.89</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>.90</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.89</td>
<td>.87</td>
<td></td>
</tr>
</tbody>
</table>
in processing but it does suggest that regression weights are no more useful than comparative information derived from them, at least under the conditions studied here.

In general, feedback was not found to have the expected positive effect on performance. Response feedback yielded performance similar to no feedback at all, while outcome feedback yielded considerably poorer performance. A principal consequence of outcome feedback was its detrimental effect on cognitive control, a detriment which was exacerbated by the addition of historical information (i.e., prior outcome data).

3. Does practice influence performance?

Learning a cognitive skill such as the one used in this research can be divided into three different stages of mental processing (Fitts, 1964). The first stage, called the cognitive stage, involves an initial encoding of the skill into a form sufficient to generate the desired behavior to some crude extent. This stage is characterized by rapid learning and sometimes verbal mediation or rehearsal while the task is being attempted. The second stage, called the associative stage, involves the "smoothing out" or perfecting of the skill performance. This stage is characterized by a slowing down of learning while gradually detecting and eliminating
errors in the initial understanding of the skill. Concomitantly, there is a dropout of verbal rehearsal. The third stage, called the autonomous stage, involves even slower learning but continued improvement in performance of the skill. The improvement in this stage continues indefinitely, or at least until the cognitive skill is mastered.

When the subjects in this experiment were presented with an optimum weighting strategy and given an opportunity to practice using it prior to the collection of expected data, it was assumed that observation of the learning process would commence at the associative stage of learning rather than at the traditional cognitive stage. Therefore, results were expected to reveal a gradual but consistent improvement in performance across blocks. The cognitive stage was expected to have ended with the 20-profile warm up session and the autonomous stage was believed to be beyond the scope of this research.

No attempt was made to counterbalance predictability over blocks and, as a result, there was a degree of confounding between these variables, as illustrated in Table 2 (maximum achievement possible). However, average predictability was approximately equal for the three most widely spaced blocks ($\bar{z} = .81, .82$, and .82 for blocks 1, 3, and 5, respectively). Consequently, analysis of
practice effects was limited to these three blocks in order to control predictability.

The results revealed that practice across blocks had the anticipated effect on achievement only in three of the five feedback groups (illustrated in Figure 8). The control, non-historical, and comparative groups showed an increase in performance with practice, but the historical and exact groups showed an increase on block 3 and a decrease on block 5, an effect which may have been due to the increased mental load or stress imposed by these feedback conditions. Both of the latter feedback types may require a more analytic mode of cognitive processing than the other three and, therefore, cause earlier fatigue than the other types. Thus the findings showed a significant effect of practice for actual achievement, \( F(2, 130) = 5.89, p < .01 \), but mean performance collapsed across the five groups did not suggest a typical learning function (0.69, 0.71, and 0.70, respectively for the three blocks). The two different trends present in actual achievement resulted in a feedback type by block interaction \( F(8, 130) = 2.19, p < .03 \).

Just as for actual achievement, actual hit-rate did not have the anticipated affect across feedback groups. As illustrated in Figure 9, performance for the control and exact groups, which was highest on block 1, declined on
block 3. Performance in the other three groups did, however, suggest a typical learning function. Thus the findings show a significant effect of practice for this measure, \( F(2, 130) = 27.51, p < .01. \) The two different trends present in the data did not result in a significant feedback type by practice blocks interaction, however, \( F(8, 130) = 1.15, p < .34. \)

Improvements in cognitive processing with practice occurred only in blocks 1 and 3. Figure 10 illustrates that knowledge averaged across groups increased from \( G = .79 \) in block 1 to \( G = .92 \) in block 3 but declined slightly in block 5 (\( G = .90 \)), \( F(2, 130) = 507.32, p < .01. \) Figure 11 illustrates a less dramatic trend for control, \( F(2, 130) = 4.95, p < .01. \) The sharp increase in knowledge between blocks 1 and 3 suggests that perhaps the cognitive stage was not completed in the warm up period as planned and that subjects were still encoding task structure information during the early blocks.

Taken together, these results suggest that practice has a significant effect on performance but the nature of the effect is specific to both the feedback type and the measure used to evaluate it. In addition, early practice influences performance by improving the subject's understanding of task structure (knowledge) rather than his application of that knowledge (control).
4. Does feedback efficacy differ as a function of predictability?

It was hypothesized that the difference between the effects of outcome and response feedback would diminish as predictability was increased. Response feedback was expected to yield high performance at all three levels of predictability tested while outcome feedback was expected to do so only when predictability was high. Therefore, the difference between outcome and response feedback performance was expected to be reduced as predictability was increased. Both actual and relative performance measures were expected to result in feedback type by predictability interactions, a result which would suggest that a task property such as predictability influences feedback efficacy.

Findings did not support this hypothesis even though a feedback type by predictability interaction was found for actual hit-rate, $F(8, 130) = 4.40, p < .01$. As illustrated on Figure 12, the difference between the response feedback groups and the outcome feedback groups was greatest when predictability was high rather than when it was low. This trend was also observed in the actual achievement measure of performance (illustrated in Figure 13) even though it was not significant, $F(8, 130) = 1.57$, .
FIGURE 12: ACTUAL HIT RATE BY PREDICTABILITY.
\( p < .14 \). These results, while unexpected, strengthen the argument that outcome feedback hinders performance.

The relative measures produced functions somewhat different from each other. Relative hit-rate (illustrated in Figure 14) depicts a sharp, negative relationship between predictability and performance, and relative achievement (Figure 15) depicts a gradual, positive one. Whatever their shape and direction, however, the effects of feedback conditions on these functions were consistent with the actual measures. Performance was degraded by outcome feedback and the expected interaction did not materialize. However, only the interaction found for relative hit-rate was significant, \( F (8, 130) = 2.42, (p < .02) \).

Apparently, the decline in relative achievement for the outcome feedback conditions (Figure 15) were not large enough to cause a feedback by predictability interaction, \( F (8, 130) = .58, p < .79 \).

These results suggest that increased predictability does not create a task environment in which response and outcome feedback will have comparable utilities. Rather, it suggests that outcome feedback is more detrimental to judgment performance than response feedback and even no feedback at all, regardless of whether performance is measured in actual or relative terms.
FIGURE 14: RELATIVE HIT RATE BY BLOCKS.
Conclusions

This research has shown that feedback efficacy is influenced by task predictability but not in the manner expected. The utility of response feedback over outcome feedback was increased rather than decreased by an increase in predictability when performance was measured in actual and relative terms. This finding is not consistent with Adelman's hypothesis that the difference between the effectiveness of outcome and cognitive feedback is negligible when task content information is available. Perhaps Adelman's investigation of congruence supported this hypothesis while the present study of predictability did not because the difference in the results of two studies implies that such effects are specific to the task properties involved. This conclusion suggests that a more systematic and thorough investigation into specific properties of judgment tasks is needed. Since feedback type was found to interact with congruence and predictability, although in different ways, it is possible that properties such as these will also interact with each other. For example, high congruence may compensate for the negative effects of low predictability or visa versa.

Since feedback efficacy was significantly influenced
by predictability in the present research, results are consistent with the Cognitive Continuum Theory even though the specific hypothesis regarding its effect was not confirmed. One might conclude, therefore, that outcome and response feedbacks are truly different points on the same continuum and consequently evoke different modes of cognition. Since outcome feedback was detrimental to performance, the CCT would suggest that it promotes more intuitive cognitive processing than response or no feedback does. Concomitantly, high predictability promotes more analytic cognitive processing.

From a more practical standpoint, the results have shown that neither response nor outcome feedback is very useful to the human judge when knowledge of task structure is available. This finding suggests that judgment performance could benefit from identifying relevant cues, determining optimum cue-criterion relationships, and perfecting presentation techniques.

An explanation for the overall inferiority of outcome feedback in the present study may lie in the fact that correct response information is never entirely consistent with a response strategy based on weighting information. That is to say, a particular judgment may be identical to that of the optimum model but, due to the stochastic nature of the judgment task, still receive
negative feedback; and a poor judgment will at times be "correct" by chance. Outcome feedback, therefore, may introduce "noise" into the cognitive system which results in lower control and a list of prior judgments may intensify this "noise." In either case, the conclusions proposed by Einhorn (1980a) and Fischhoff (1975) that outcome feedback is irrelevant for correcting inappropriate response strategies is supported by this research.

The changes in the cognitive processing measures used in this investigation (knowledge and control) suggest that processing limitations on performance cannot be eliminated by increasing task predictability. The nature of these limitations support the use of a judgment aid such "bootstrapping," the procedure by which the judge is replaced by his model (Dawes & Corrigan, 1974). Such an approach would eliminate to some extent the processing inconsistency caused by poor control, particularly in low predictability environments.
Reference Notes


References


Brehmer, B. Effects of cue validity and task predictability on interpersonal learning of linear inference tasks.
Organizational Behavior and Human Performance, 1974, 12, 17-29. (a)

Brehmer, B. Hypothesis about relations between scaled variables in the learning of probabilistic inference task. Organizational Behavior and Human Performance, 1974, 11, 1-27. (b)


Dees, V., & Grindley, G. C. The effect of knowledge of
results on learning and performance. The direction of the error in very simple skills. Quarterly Journal of Experimental Psychology, 1951, 1, 36-42.


Einhorn, H. J. Overconfidence in judgment. New Directions for Methodology of Social and Behavioral Science, 1980, 4, 1-16. (b)


Miller, P. McC. Do labels mislead? A multiple-cue study within the framework of Brunswik's probabilistic functionalism. *Organizational Behavior and Human Performance*, 1971, 6, 480-500.


Norman, K. L. Rule learning in a stimulus integration task. *Journal of Experimental Psychology, 1974, 103,* 941-947. (b)

Norman, K. L. Effects of feedback on the weights and subjective values in an information integration model. *Organizational Behavior and Human Performance,* 17, 367-387.


Smode, A. F. Learning and performance in a tracking task under two levels of achievement information feedback. *Journal of Experimental Psychology,* 1958, 56, 297-304.


