ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
Vol. 70, No. 1, April, pp. 41–64, 1997
ARTICLE NO. OB972692

Looking and Weighting in Judgment and Choice

Douglas H. Wedell and Stuart M. Senter

University of South Carolina

A sampling model was proposed in which the weight given to a piece of information corresponds to the amount of sampling of that information in either a continuous, discrete or strategic manner. These three sampling processes were related to process tracing measures of initial and additional time per acquisition and frequency of acquisition. The applicability of the sampling model was tested in three experiments in which students uncovered information corresponding to verbal and math aptitude scores of hypothetical applicants and either judged the likelihood of success in a designated major or chose which of a pair of applicants was more likely to succeed in the major. Task focus was manipulated by altering the designated major. In Experiment 1, analysis of judgment data demonstrated large effects of task focus on the weighting of verbal and math scores and corresponding increases in number of acquisitions and time per acquisition on the information receiving more weight. In Experiments 2 and 3, analyses of choice proportions revealed effects of task focus on weight and bias parameters. Looking data in choice provided strong support for two of the stages of processing described by Russo and Leclerc (1994). Initial looks reflected orientation and screening functions and additional looks reflected more evaluative processes. Experiment 3 also explored similarities and differences among groups of participants who were classified as following different identifiable choice strategies.

Process tracing has proven a useful tool in helping decision scientists understand decision processes (for reviews see Ford, Schmitt, Schechtman, Hults, & Doherty, 1989; Payne, Bettman & Johnson, 1992). The term process tracing can be applied to a wide variety of techniques that include online and retrospective verbal protocols, information board paradigms, and eye movement monitoring. Our focus is on the typical information board process tracing task used in decision making, in which information is masked on a computer screen and uncovered by moving a cursor into the masked area. The data record includes the option chosen or judgment rendered, the sequence of information accessed, and the amount of time looking at information. A major goal of process tracing studies has been to understand the different strategies that decision makers use in choosing an option. Strategies differ in (a) how information is accessed (dimensionwise or alternativewise) and (b) how information is valued and integrated (qualitatively or quantitatively). Process tracing measures, such as those describing the relative amount of alternativewise processing, the completeness of information search, and the order of accessing information have been used to infer choice strategies (Böckenholt, Albert, Aschenbrenner, & Schmalhofer, 1991; Payne, Bettman, & Johnson, 1988; Russo & Dosher, 1983; Schkade & Johnson, 1982; Svenson, 1979). For example, the tendency to look across alternatives on a single dimension and then make a choice would imply a lexicographic strategy (Tversky, 1969). At the other extreme, an exhaustive and time consuming search of information for a given alternative before proceeding to the next alternative is consistent with a weighted additive strategy (Payne *et al.*, 1988).

The use of a variety of different strategies in choice may be one reason that results from choice and judgment tasks do not always agree. These differences have been documented in the preference reversal phenomenon (Goldstein & Einhorn, 1987; Lichtenstein & Slovic, 1971, 1973; Slovic & Lichtenstein, 1983) and discrepancies between choice and matching (Tversky, Sattath, & Slovic, 1988). In a judgment task, information is typically presented one alternative at a time, and the judge must give an overall evaluation of that alternative. Thus, the judgment task is likely to induce a weighted additive strategy based on alternativewise processing of information. The results from numerous judgment experiments demonstrate the good fit of weighted additive models to a wide variety of judgment tasks (cf., Anderson, 1981). On the other hand, many choice strategies are noncompensatory in nature and may be performed without ever forming an overall integrated

This research was conducted under NSF Grant SBR-9319520.

Address correspondence and reprint requests to Douglas H. Wedell, Department of Psychology, University of South Carolina, Columbia, SC 29208. Internet address: Wedell@sc.edu.

0749-5978/97 $25.00

impression of the alternative that is chosen. For example, the satisficing and elimination by aspects strategies require only that the chosen alternative exceed some threshold value on each attribute. The lexicographic strategy requires only that the chosen alternative has the highest value on the most important dimension. The majority of confirming dimensions strategy requires only that the dimensional values of the chosen alternative exceed other alternatives on a majority of dimensions. None of these strategies requires the quantitative valuation and integration of information entailed by the weighted additive strategy. In contrast to the alternativewise processing typical of judgment, process tracing studies have demonstrated that decision makers overwhelmingly use noncompensatory dimensionwise strategies in choice, especially when the number of alternatives or attributes is large (Ford *et al.*, 1989; Payne, 1982).

The focus of the present experiments was on the relationship between looking time and weighting of information in judgment and pairwise choice. These experiments were guided by the research question of how a change in task focus affects looking behavior. A common assumption of process tracing approaches is that weight is reflected in looking time measures. Although often assumed, there have only been a few investigations that have provided some validation for this assumption (Fiske, 1980; Schkade & Johnson, 1982). One theoretical justification for a correspondence between weight and looking time is that weight may reflect the attention given to a stimulus (Fiske, 1980). However, at this time, there is simply no clear agreement on what cognitive processes correspond to the weighting parameters found in judgment and choice models. In structural models, weight describes the relative influence of a piece of information, but these formal models do not specify how weights operate within a cognitive processing system. In this article we develop a sampling model of weighting that provides theoretical linkage between online process measures and the weighting parameters derived from models of judgment and choice outcomes. The sampling model derives from work in choice and absolute and comparative judgment in which latencies are assumed to depend largely on the number of samples of the stimulus information the judge gathers before making a response (Busemeyer & Townsend, 1993; Link, 1992; Petrusic, 1992). The simple notion that each sampling of information requires a fixed unit of time yields predictions of effects of varying discriminability, symbolic distance, and speed–accuracy instructions. The sampling model presented here differs from previous models primarily because its focus is on the weighting process (rather than stimulus valuing) and it seeks to predict looking times at specific pieces of information rather than judgment latencies.

A SAMPLING MODEL OF WEIGHT

The basic tenet of the sampling model presented here is that the weight given to a piece of information corresponds to the degree to which that information has been sampled: the greater the weight, the greater the sampling. We further assume that the sampling of information is reflected in the looking behavior of the decision maker. We will distinguish three types of sampling behaviors and link each of these to a corresponding process measure. However, before doing so, we first present an example of how weighting and sampling may correspond.

For illustrative purposes, consider a simple weighted-additive model of how a person described along two dimensions is judged. Following Anderson (1981), the integrated impression of person *j* described along dimensions 1 and 2 may be represented within a constant weight averaging model as follows:

$$\mathrm{I}_j = w_0\mathrm{S}_0 + w_1\mathrm{S}_{j1} + w_2\mathrm{S}_{j2}, \tag{1}$$

with the weights (w_0, w_1, w_2) constrained to sum to 1.0. Greater weight for a dimension means that stimulus values along that dimension exert greater influence on the overall judgment than stimulus values on other dimensions. One way to conceive of the judgment process represented by Equation (1) is as an anchoring and adjustment process (Lopes, 1981), with weight influencing the degree of adjustment resulting from a given piece of information. Within the sampling framework, increased sampling leads to increased adjustment toward the sampled value and hence greater weight of that piece of information.

This process is illustrated in the examples shown in Fig. 1. The top panel represents a situation in which stimulus values 1 and 2 are sampled equally. The starting point of the process is the initial value, S_0, which corresponds to a 5 on a 9-point internal judgment scale. The adjustment toward each successively encountered value is simply $1/k$, where k represents the number of samples taken at that moment. Thus, when S_1 is sampled, the internal judgment is adjusted $\frac{1}{2}$ of the way from the current value to the value of S_1. On each successive sample, the adjustment toward the sampled value reduces proportionally as a function of the number of samples. Because each value (S_1 and S_2) is sampled twice in the top panel, the two contribute equally to the overall judgment and hence are given equal weight.

The bottom panel of Fig. 1 illustrates oversampling

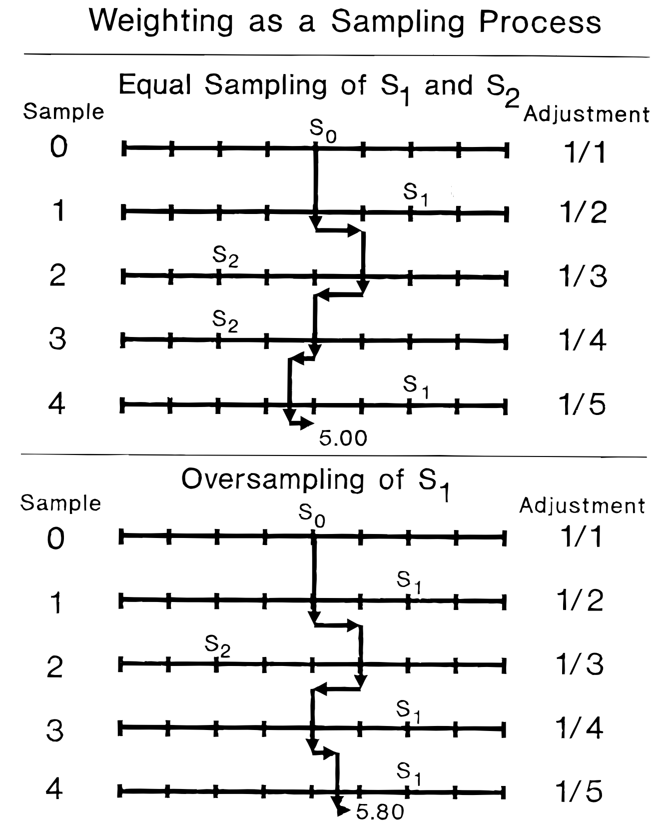

FIG. 1. Schematic illustration of how weighting corresponds to sampling of stimulus values within an anchoring and adjustment framework. Adjustment from the current position to the position of the sampled value is $1/k$, where k is the number of samples at that point in the sampling process. Weight of a value corresponds to the number of samples of that value divided by the total number of samples.

of S_1, corresponding to greater weight given to S_1. Because S_1 is sampled more than S_2, there is greater adjustment toward its value. The greater adjustment toward S_1 is modeled within the structural model as greater weight given to S_1. Specifically, the weight of a stimulus is simply the number of times it was sampled divided by the total number of samples. Thus, the inferred weights for the top panel of Fig. 1 are $w_0 = 1/5$, $w_1 = 2/5$, and $w_2 = 2/5$. The inferred weights in the bottom panel are $w_0 = 1/5$, $w_1 = 3/5$, and $w_2 = 1/5$. Within this framework then, the inferred weights from the judgment task reflect the relative frequency of sampling of information, which in turn leads to a greater adjustment of the internal judgment towards the sampled value. The example illustrated in Fig. 1 represents one basic way in which increased sampling can lead to greater weight. Below we draw distinctions among three different ways in which the sampling process may operate.

Continual Sampling

One way in which weighting may operate is as continual sampling within a given look. In other words, when an important piece of information is encountered, greater importance or weight leads to increased sampling of that information and greater adjustment of internal judgments towards that value before moving on to another piece of information. This view of weighting predicts that greater weight given a piece of information should be reflected in greater time per acquisition (TPAQ) for that piece of information.

Not all views of the weighting process predict corresponding positive shifts in TPAQ. For example, one may look at one piece of information while processing a different piece of information. In this view, TPAQ may well be unrelated to weighting. Looking TPAQ may simply correspond to reading the information, whereas processing the information may occur while one looks at other information or after all the information has been examined. Attentional accounts of the weighting process may even reverse the predicted positive relationship between looking TPAQ and weight. Across many cognitive tasks, increasing the attentional resources aimed at a piece of information facilitates or speeds up processing of that information (Posner & Snyder, 1975). If increasing weight means increasing attention, then greater processing of the information may not be reflected in greater looking TPAQ because of the speed up in processing when attentional resources are focused on the information. Indeed, TPAQ may actually be reduced for high weight information due to facilitative effects of attentional focus. (For an alternative view of the relationship between attention and weight, see Fiske, 1980.)

Discrete Sampling

Another way in which a sampling model of weight might operate is at the level of the number of discrete looks at information rather than the looking TPAQ. Thus, for example, weight may not be reflected in the amount of time for a given look, but instead it may be reflected in the repeated reaccessing of information deemed most important. Once again, this greater sampling may lead to greater adjustment of the internal judgment of the stimulus toward the value of the information being sampled. This view is consistent with evidence from the literature on eye movement monitoring of textual and nontextual displays that information that is most important to the goals of the processors will accessed more frequently than less important information (Duffy and Rayner, 1990; Hegarty, 1992; Rayner & Morris, 1990).

However, it is again possible to develop alternative views of the weighting process. One might argue that information that is most important requires the least resampling because it has a privileged place in working memory. Such a viewpoint could lead to the opposite prediction that greater looking time would result in a reduction of reaccesing of information.

Strategic Sampling

Participants do not always look at all of the information provided. This is particularly true in choice tasks in which noncompensatory strategies may be used. Consider a lexicographic strategy in which information is first examined and compared on the most important dimension. One only proceeds to the next dimension if there is no clear winner. This sense of weight reflects a bias in selection of information rather than adjustment toward stimulus values. It can affect sampling in two ways. First, if information locations are known to the participants, then the dimension with the greatest weight should be sampled first. Hence, weight affects the priority of sampling. Second, dimensions of less importance will be less likely to be sampled even once, because a decision may have been reached after sampling the more important dimension. For example, stopping rules in the lexicographic and elimination by aspects strategies may lead to only a single most important dimension being sampled. This type of strategic sampling predicts that the number of times information will be accessed at least once will be higher for dimensions given greater weight.

There is considerable support for strategic sampling in choice. The most important attribute is more likely to be sampled first and less important attributes are less likely to be sampled even once (Payne et al, 1988; Russo & Rosen, 1975; Russo & Dosher, 1983). It should be noted that the sense of the term weight used in strategic sampling is rather different that that used in continuous and discrete sampling. In those two sampling processes, weight corresponds to an incremental operation on values. However, the meaning of weight in strategic sampling appears to be more along the lines of an all or none process determining whether an attribute value will be sampled at all.

Summary of Hypotheses

To summarize, these sampling models of weighting behavior lead to three specific hypotheses.

H1: If sampling occurs in a continual fashion within a look, then greater dimensional weight should lead to greater looking TPAQ for that dimension.

H2: If sampling reflects discrete accessing of information, then greater dimensional weight should lead to greater frequency of additional accesses to information on that dimension.

H3: If participants engage in strategic sampling, then greater dimensional weight should lead to greater frequency of initial access to information on that dimension.

Note that all three of these hypotheses lead to the prediction of increased looking time with increased weight, but they differ in how the link to looking time is achieved. Of the three hypotheses, H3 is the most firmly established. However, previous research relating looking time to weight has typically examined this relationship using between-subject comparisons. In the research described here, we develop a more powerful test of the relationship between weighting and looking by manipulating task focus within-subjects. This manipulation allows us to compare looking behavior of the same subject for the same materials under two different weighting conditions and hence provides a very powerful test of the hypothesized relationships between looking and weighting.

OVERVIEW OF EXPERIMENTS

We report three experiments conducted to the test the above hypotheses. These experiments share several common features. First, we used very simple stimulus materials. In all three cases, the information presented to participants was simply verbal and math aptitude scores of hypothetical students seeking admission into a university. Participants were to imagine they were part of an admissions committee that would consider applications. They were told that although more information would be considered in making an admission, their job was to make an initial determination of likelihood of success based on these two aptitude scores. In the judgment task of Experiment 1, they were to make estimates of the likelihood of the student succeeding in the specified major. In the choice tasks of Experiments 2 and 3, they simply selected which of the two students presented on the screen had the higher likelihood of success. In all three experiments, the task was set up so that score information was hidden in unlabeled boxes on the computer screen. Thus, the participants did not know which boxes contained verbal scores and which contained math scores on a given trial. We used this procedure because we wanted to encourage participants to look at each type of score information at least once in order to study their weighting behavior.

We manipulated task focus in order to alter participants' weighting of verbal and math scores. In Experiments 1 and 2, the scores were attributed to students who were applying either to an English or an engineering major. Participants were reminded that the

English major included many courses with difficult reading and extensive writing assignments. They were told that the engineering major included many courses with advanced mathematics. In Experiment 3, we sought to reduce the differences in the task focus manipulation by using more similar majors, sociology and economics. While the sociology major emphasized verbal abilities, it did include some math. Conversely, the economics major emphasized mathematical skills, but it also included demands on verbal skills. In all cases, major was manipulated within-subjects so that participants rated or chose between hypothetical students in a first block of trials referring to one major and then in a second block of trials referring to the other major. Order in which majors were considered was counterbalanced across participants.

Two classes of dependent variables were used. The first of these was a response outcome measure: either ratings or choices. These responses were used to infer the relative weighting of verbal and math scores in each major. The second class of dependent variables was process tracing measures. These measures were broken down into initial and additional looks in the following ways.

- *Initial looking TPAQ* was the time (in milliseconds) spent on an initial look at a piece of information, given there was a look. If the information was never viewed, then no value was entered for initial TPAQ on that trial.
- *Initial frequency of access* was a 1-0 variable indicating whether or not the information was accessed at least once.
- *Additional looking TPAQ* was the mean time (in milliseconds) spent on each additional look at a piece of information. If the information was viewed only once, then no value was entered for additional TPAQ.
- *Additional frequency of access* was the total number of accesses to a piece of information after the information had been accessed at least once.

Our strategy in these three experiments was to examine rigorously how looking and weighting correspond. By structurally modeling outcome data and examining the time course of processing, we hoped to develop a more complete picture of how weighting parameters of structural models correspond to observable process measures.

The breakdown of looking behavior into initial and additional looking measures also allowed us to examine the stagelike quality of choice. Russo and Leclerc (1994) have postulated three stages in the choice process. These are (a) orientation and screening, (b) evaluation, and (c) verification. The orientation stage corresponds to the initial looking at alternatives in order to sample and screen the available information. Evaluation includes extensive comparison among the alternatives. The final stage, verification, includes examination of previously unexamined alternatives in order to verify that the tentatively selected alternative is better than all other alternatives. Russo and Leclerc (1994) applied this framework to choices among a large set of consumer products.

We examined the extent to which this framework applies to judgments and choices when there are very few pieces of relevant information. We hypothesized that the judgment task would bypass the initial screening or orientation stage, because there is no need to eliminate alternatives from consideration. Similarly there is no need for a verification stage, simply because there are no other competing alternatives. Thus, we predicted that the evaluative process would occur within the very first looks and extend throughout the examination of the stimulus. The two-alternative, two-attribute choice paradigm of Experiments 2 and 3 provided a greater opportunity for participants to employ the initial orientation and screening processes. Evidence for such processes would involve initial looking behavior differing in important ways from additional looking behavior or looking behavior within the judgment task.

EXPERIMENT 1: JUDGMENTS OF ENGINEERING AND ENGLISH MAJORS

The purpose of Experiment 1 was to determine the correspondence between looking measures and weight in a judgment task. The establishment of a strong positive relationship would provide some validity of the assumption that looking time reflects weight. A positive association between weight and looking time has been demonstrated in some choice tasks (Payne *et al.*, 1988; Wedell, 1993) and judgment tasks (Fiske, 1980; Schkade & Johnson, 1982). However, those studies did not examine within-subject changes in weighting when participants were examining the same materials. This methodology should provide a more powerful method for examining correspondences between looking and weighting.

In Experiment 1, judges viewed hypothetical test information corresponding to verbal or math aptitude scores and predicted how successful the prospective student would be in either an engineering or English major. Judges moved a mouse cursor into boxes to unmask score information prior to indicating a judgment on a 9-point scale. The judgment data was analyzed using a constant weight averaging model described by Eq. (1). Modeling the weights allowed us to verify that the manipulation of task focus had the predicted effect on the relative weighting of score information. It also allowed us to derive weights for each participant and determine whether differences in weights across participants corresponded to differences in process measures.

Method

Participants and Design

Participants were 42 students from a southern university, who received course credit for their participation. The basic design consisted of a $2 \times 5 \times 5$ factorial combination of rating task (predict success in engineering or English major), verbal score (five levels), and math score (five levels). The order in which the two tasks were presented was counterbalanced so that half of the participants judged success in English in the first block of trials and half judged success in engineering in the first block of trials. Presentation of pairs within each block of trials was randomized. The dependent variables were the rating of predicted success on a 9-point scale and the looking behavior measures. After judging the paired scores from the two sets, participants judged the likelihood of success in college for each level of the verbal and math scores that appeared in the study.

Materials and Apparatus

All instructions and stimuli were presented on microcomputers, and responses were collected via the keyboard and mouse. For the engineering judgment task, the five verbal scores ranged from 350 to 670 in increments of 80 and the five math scores varied from 370 to 690 in increments of 80. For the English judgment task, the five verbal scores ranged from 360 to 680 in increments of 80 and the five math scores ranged from 340 to 660 in increments of 80. In each major, the scores were combined to form 25 pairings. Altogether, 20 different score values were used, 5 math scores and 5 verbal scores for each major. The average of math scores was 20 points higher than the average of verbal scores in the engineering rating task and the average of verbal scores was 20 points higher than the average of math scores in the English rating task.

Procedure

Participants were instructed that the task concerned how people make evaluations of others based on score information. They were told to imagine they were part of an admissions board at a university and their task was to predict success in a given major based on aptitude scores. Judges were told the basic characteristics of these scores, including that the mean of scores was 500, the standard deviation was 100, less than 3% of scores fell below 300, and less than 3% of scores rose above 700. They were told that the committee would use more information than just the aptitude scores, but that their task was to predict success in the major based solely on these scores. Ratings of success were made on a 9-point scale, with end points labeled “very unlikely to succeed” and “very likely to succeed,” respectively. They were told that the two scores would be hidden behind boxes on the screen and that they should open the boxes by moving the mouse pointer into the box. When the mouse pointer moved into the box, the score was exposed, and when the pointer left the box, the score was hidden again. To register their judgments, participants moved the mouse pointer to the number corresponding to the rating and clicked the mouse button. To confirm the rating, they moved the mouse pointer into a box labeled “ok” and clicked. Prior to clicking the “ok” box, they could change the rating by simply selecting another number.

Before each set of ratings, participants were given a practice trial to get acquainted with using the mouse to open boxes and record ratings. The experimental trials for composite ratings were presented in two blocks of 25 trials, corresponding to the two sets of 25 score combinations. There was a one minute rest period between presentation of each set. Presentation of score pairs was randomized within each block, and which score was presented in the left or right box on the screen was also randomized, so that the judges did not know prior to opening a box which box contained the math score and which contained the verbal score. Score labels were hidden in the boxes to increase the likelihood that participants would look at both scores rather than adopt a strategy in which one score could be consistently skipped.

Before judging success in the engineering major, participants were told that the major consisted of many courses involving highly complex math. Before judging success in the English major, participants were told that the major consisted of many courses involving reading difficult material. After two sets of composite ratings, participants were asked to rate success of individuals described by a single math or verbal score. The 20 scores (10 from each of the two sets) were presented in random order. Once again, the scores were hidden in boxes that were opened with the mouse pointer. Participants were told to predict success in a general studies major that used math and verbal scores equally.

Results

Single Score Ratings

Figure 2 presents the mean ratings of the single scores used in the two tasks. These data provide a manipulation check for the experimental design. In general, the mean ratings reflected the five roughly evenly spaced levels of performance on the two score dimensions. As reflected in the design, the math scores received slightly higher ratings than the corresponding verbal scores in the engineering condition, and the opposite was true for the English condition. The spacing of the five levels of scores was approximately linear. Separate repeated measures ANOVAs conducted on math and verbal scores each reflected large main effects of score level ($p < .001$). In each case, nearly all the variance was carried in the linear component (99.6% for math and 99.4% for verbal), although both sets of scores resulted in significant higher order polynomial trends ($p < .05$).

Combined Score Predictions

Figure 3 presents the mean ratings for the combined scores for the two tasks, along with the fit of the constant weight averaging model. The change in the relative weighting of math and verbal scores is clearly evident in the pattern of data. In the engineering judgment task, the slopes of the rating functions were very shallow, reflecting the low weight given verbal scores. In the English judgment task, the slopes of the rating functions were very steep, reflecting the high weight

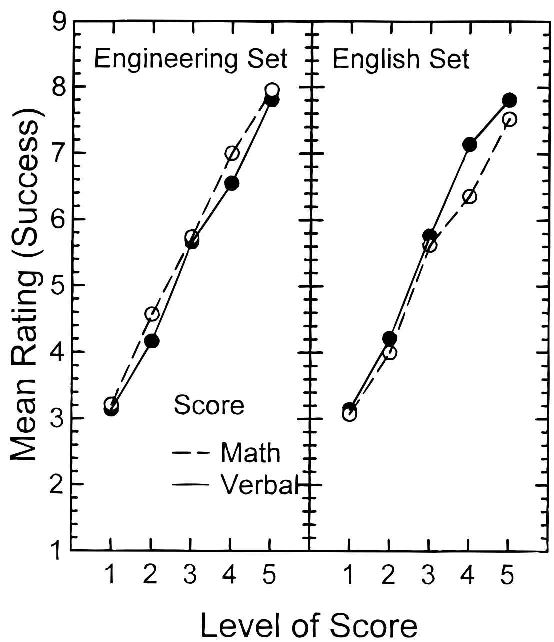

FIG. 2. Single score ratings of predicted success (Experiment 1).

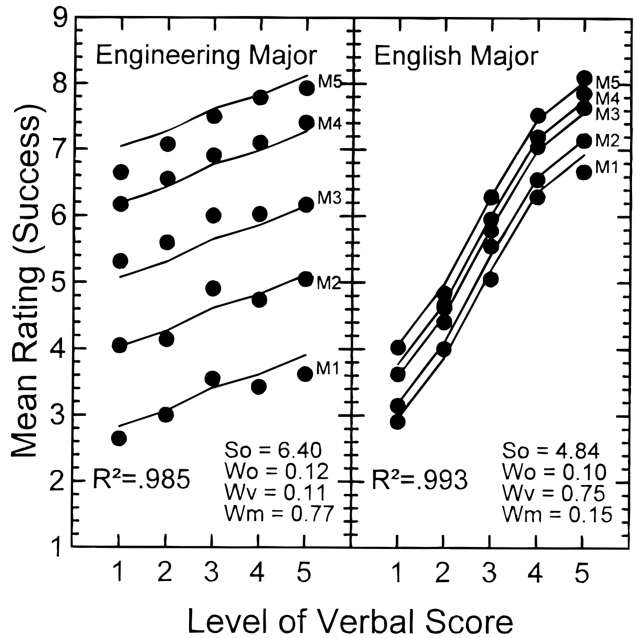

FIG. 3. Fit of the constant weight averaging model to the mean ratings for combined scores reflects the strong shift in weights across majors (Experiment 1). Note that M1–M5 are the five levels of math scores, S_0 is the initial impression, w_0 is the weight of the initial impression, and w_V and w_M are weights of verbal and math scores, respectively.

given to verbal scores. The opposite weighting pattern is evident for math scores, with math scores having much greater influence over judgments of success for the engineering judgment task than for the English judgment task.

A $2 \times 5 \times 5$ repeated-measures ANOVA was conducted on the ratings of the 25 combined scores in the two conditions. The highly significant main effects of verbal and math scores (Fs > 250) simply reflected the higher ratings associated with higher scores. More important to the question of change in dimensional weighting, the Task $\times$ Verbal Score and Task $\times$ Math Score interactions were both highly significant, Fs > 50. Thus the task focus manipulation successfully shifted weighting of the score information in the expected direction. The Verbal Score $\times$ Math Score interaction was not significant, supporting the applicability of a constant weight averaging model which predicts parallel effects at each score level.[1]

[1] A lack of an interaction is typically taken as support for the constant weight averaging model. However, the greater influence of negative information on judgment is well documented (Birnbaum, 1974; Skowronski & Carlston, 1989) and was captured in a significant linear x linear component of the interaction. This effect could be modeled by a differential weight, configural weight or geometric averaging model. However, because the effect was small and because the generalization of the constant weight model to choice is more

The constant weight averaging model provided a good fit to the data, explaining approximately 99% of the variance in mean ratings with only three fitted parameters.[2] In fitting the data, the weighting of verbal scores versus math scores was .11 to .77 in the engineering condition and .75 to .15 in the English condition. These weights reflect the large shift in the relative influence of verbal and math scores in the two tasks.

Looking Measures

The critical questions addressed in Experiment 1 concern the relationship between looking measures and dimensional weight. The rating data clearly indicated a crossover interaction in dimensional weighting, with greater weight given to verbal scores than to math scores when judging English majors, but the reverse was true when judging engineering majors. If there is a close link between looking measures and weight, then the same type of crossover interaction should be observed for the looking measures.

Figure 4 presents the TPAQ and frequency of access looking measures for verbal and math scores for both initial and additional looks. All four panels reflect the predicted Major × Dimension interaction, providing support for all three sampling hypotheses. The top panels of Fig. 4 describe the initial and additional looking TPAQ, which reflect how long information was viewed on a given access. Consistent with the continual sampling model (Hypothesis 1), TPAQ showed the predicted Major × Dimension interaction: There was greater initial and additional TPAQ for information that received greater weight.

Three participants who failed to look a second time at information within one cell of the design were eliminated from the 2 × 2 × 2 repeated-measures ANOVA conducted on TPAQ. The main effect of block reflected slightly reduced TPAQ for additional looks, $F(1, 38) = 6.4$, $p < .05$. The only other significant effect was the Major × Dimension interaction, $F(1, 38) = 6.5$, $p < .05$. The three-way interaction was not significant. Planned comparisons revealed a significant Major × Dimension interaction for initial TPAQ but not for additional TPAQ (although this effect approached significance, $F(1, 38)$

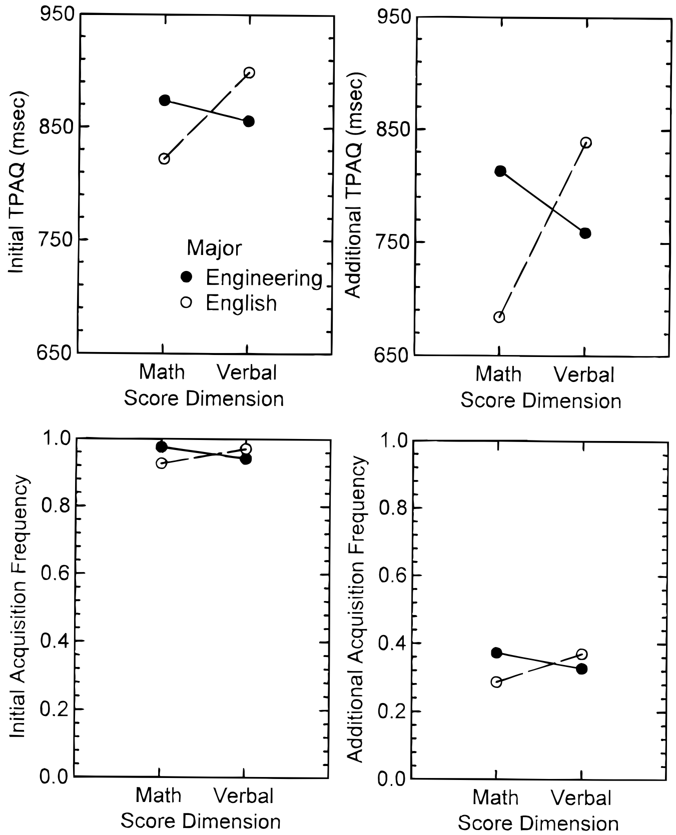

FIG. 4. Initial and additional time per acquisition (TPAQ) and frequency of access as a function of major and score (Experiment 1). Interaction patterns are significant in each panel, providing support for continuous, discrete and strategic sampling processes.

$= 4.0$, $p < .06$. Thus, when participants looked at information, they tended to look longer at information to which they gave greater weight.

Although each piece of information was initially accessed more than 90% of the time, the information that was most relevant to the task was skipped less often, as shown in the bottom left-hand panel of Fig. 4. This effect is consistent with the strategic sampling model (Hypothesis 3). After viewing information once, that information was accessed again on average only about one third of the time. The information receiving less weight was also less likely to be accessed again than information receiving more weight. This effect was consistent with the discrete sampling model of weighting (Hypothesis 2).

A 2 × 2 × 2 repeated-measures ANOVA conducted on the frequency data revealed two significant effects. The main effect of stage reflected the greatly reduced frequency of access for additional looks. More importantly, the Major × Dimension interaction was significant, $F(1, 41) = 20.9$, $p < .001$. There was no Stage × Major × Dimension interaction. Planned comparisons revealed significant Major × Dimension interactions for both initial access frequency and additional access frequency

straightforward, we chose to fit the data using the constant weight averaging model.

[2] The constant weight model was fit using a least squares iterative nonlinear regression technique. The single score ratings were used to infer scale values. The use of the single score ratings also allowed us unambiguously to fit the initial impression S_0 and its weighting w_0. Because scale values were determined by the single stimulus ratings, the paired rating data shown in each panel of Fig. 3 was modeled with only three free parameters (S_0, w_0, and w_V), with the constraint that $w_M = 1 - (w_0 + w_V)$.

($p < .01$). Thus, judges looked more often initially and on additional looks at the information receiving greater weight.

Individual Differences

Another way to look at the relationship between looking time and weight is to see whether individual differences in the two measures corresponded. Each judge's ratings were fit to Equation 1 using least squares iterated nonlinear regression to estimate weights for math and verbal scores in the two conditions. A score representing changes in weight was calculated by subtracting the weights of the nonfocal dimensions from the weights of the focal dimensions ($w_{\text{Math,Engineering}} + w_{\text{Verbal,English}} - w_{\text{Verbal,Engineering}} - w_{\text{Math,English}}$). Similarly, the relative portion of total time spent on verbal and math scores in the two conditions was calculated for each judge ($\text{Time}p_{\text{Math,Engineering}} + \text{Time}p_{\text{Verbal,English}} - \text{Time}p_{\text{Verbal,Engineering}} - \text{Time}p_{\text{Math,English}}$), with Time$p$ reflecting the time on a piece of information divided by the total time spent on information on a given trial. Proportion of time was used because there were very large individual differences in the total time participants spent looking at information on a given trial. The correlation between these two scores was significant, $r = .50$, $p < .001$, indicating that the shift in the relative weighting of verbal scores across the two conditions was accompanied by a corresponding shift in the proportion of total time spent on verbal scores.

Discussion

Experiment 1 provided clear evidence that people tend to look where they weight. Participants showed much greater sensitivity to differences in verbal scores when judging success in the English major than when judging success in the engineering major. This differential sensitivity was reflected in the weights assigned these dimensions within the constant weight averaging model. Looking behavior was broken down into initial and additional frequency of access and TPAQ. The assumption that weight is reflected in increased looking time predicted that these measures would be greater for the verbal scores in the English judgment task and the math scores in the engineering judgment task. The predicted relationship was obtained for both initial and additional looks.

The obtained Major × Dimension interaction on TPAQ supports a continual sampling model in which the judge examines information and adjusts an internal judgment toward the value of that information. Greater weight corresponds to greater sampling and adjustment as reflected in TPAQ. This type of adjustment process appears to occur the initial time the information is accessed as well as on subsequent accessing of the information. The Major × Dimension interaction on the additional frequency measure provided support for the discrete sampling model. Information that receives greater weight is also reaccessed more frequently. Both the effects on continual and discrete sampling parallel findings from the literature on eye fixations in reading, where information that is most critical to understanding the content or syntactic structure of the passage is fixated on longer and returned to more often (Duffy & Rayner, 1990; Rayner & Morris, 1990).

Although information was rarely skipped altogether, the significant Major × Dimension interaction on initial frequencies indicated that when information was skipped, it tended to be on the dimension receiving little weight. The failure to look at a piece of information even once supports the strategic sampling model and presumably corresponds to the use of a noncompensatory judgment strategy in which information on the less important dimension cannot compensate for an extreme value on the more important dimension. Thus, even in a single stimulus judgment task there may be some small tendency to employ noncompensatory strategies.

In relationship to Russo and Leclerc's (1994) three-stage model, it appears that looking behavior in this type of simple judgment task is predominantly within the evaluative stage of processing. The basic finding that initial and additional TPAQ show the same differential processing of dimensions provides no evidence that initial looks include a screening or orientation stage different from the evaluative focus of later looks. The screening phase, however, may not be totally absent, because there is some evidence for strategic sampling. Thus, on a minority of trials, less relevant information may be skipped altogether as a result of an initial screening process. Nevertheless, the results are clearly consistent with minimal use of an initial screening process, as one might expect in single stimulus judgment.

The above conclusions were based on within-subjects comparisons. The correlational evidence indicated that differences in relative looking times appeared to capture individual differences in the degree to which judges shifted the weighting of information across tasks. Judges clearly differed in the degree to which their weighting of scores changed across tasks. Consistent with the link between looking time and weight, these differences in weight correlated significantly with corresponding changes in the proportion of total looking time spent on the two types of information. This correlation adds further support to the hypothesis that looking time reflects a weighting component.

EXPERIMENT 2: CHOICES WITHIN ENGINEERING OR ENGLISH MAJORS

The results of Experiment 1 provided strong evidence for a clear relationship between looking time and weight in a judgment task. It also provided support for all three types of sampling processes operating in a judgment task, although strategic sampling was minimal. To what extent do these results generalize to choice? Because information processing strategies may differ markedly in judgment and choice, there is no guarantee that these results will generalize to choice. For example, if decision makers use a lexicographic strategy in which they simply select the highest value on the most important dimension, there may be no processes corresponding to on-line weighting of information that occurs during viewing. Use of this strategy would produce evidence for strategic sampling, but not necessarily any evidence for continual or discrete sampling. Indeed there is ample evidence that strategic sampling occurs in complex choice tasks (Payne, et al., 1988; Russo & Dosher, 1982). This type of differential weighting of attributes within a noncompensatory strategy appears to reflect weights operating on a selection mechanism that determines which attributes to process and which to ignore. It is not clear whether weighting processes that operate on values may be occurring in conjunction with weights operating on initial selection of information. The latter process has been amply supported by empirical evidence from choice tasks, but the former has not.

A relationship between TPAQ and weighting would seem to require that decision makers engage in an on-line weighting process that operates on the values attributed to each piece of information. This should occur if participants use a weighted additive strategy, which can be seen as a direct generalization of the judgment strategy. For example, the weighted additive model may be generalized to pairwise choice by modeling choice as a function of the difference of two impressions:

$$\mathrm{I}_j - \mathrm{I}_k = (w_0\mathrm{S}_0 + w_1\mathrm{S}_{j1} + w_2\mathrm{S}_{j2}) - (w_0\mathrm{S}_0 + w_1\mathrm{S}_{k1} + w_2\mathrm{S}_{k2}). \quad (2)$$

This type of difference operation is not limited to alternativewise processing of information. Tversky (1969) pointed out that from a dimensionwise perspective, Eq. (2) may be represented as an additive difference model. According to this model, differences in values on each dimension are assessed and these differences are then weighted. Thus, the additive difference representation describes the basis of choice as the weighted sum of differences on each dimension:

$$\mathrm{I}_j - \mathrm{I}_k = w_1(\mathrm{S}_{j1} - \mathrm{S}_{k1}) + w_2(\mathrm{S}_{j2} - \mathrm{S}_{k2}). \quad (3)$$

Equations (2) and (3) are equivalent and allow us to predict differences in choices based on the weighted differences of dimensional values. To predict choice proportions, we must describe the response function relating differences in impressions to choice proportions. For simplicity, we will use a logistic transformation as described by Luce's (1959) choice rule. Thus, the proportion choosing j over k based on the weighted additive or the weighted difference strategies is given by

$$\mathrm{p}_{jk} = 1/(1 + \exp(-b(w_1(\mathrm{S}_{j1} - \mathrm{S}_{k1}) + w_2(\mathrm{S}_{j2} - \mathrm{S}_{k2}))), \quad (4)$$

where b represents a scaling or discriminability constant. When the weighted differences sum to 0.0, the predicted proportion will be 0.5. If decision makers are following either of the weighting strategies described in Eqs. (2) or (3), then choice proportions should be sensitive to the magnitude of differences along dimensions 1 and 2 as described in Eq. (4).

On the other hand, consider those persons who use a simple lexicographic strategy consisting of choosing the alternative that is highest on the most important dimension. For these individuals, choice proportions will not be sensitive to the magnitude of differences along dimensions 1 and 2. Instead, the choice proportion may be predicted by a simple response bias parameter model,

$$\mathrm{p}_{jk} = 1/(1 + \exp(-a)), \quad (5)$$

with the sign of a being positive when j has a higher value than k on the most important dimension and negative when j has a lower value than k on the most important dimension.

Equations (4) and (5) correspond to relatively pure strategy models. However, it is reasonable to assume that within a sample, or within an individual, there may be a mixture of strategies used. For example, consider an application of the elimination by aspects strategy in which both alternatives are acceptable on each dimension. The individual may then turn to a weighted additive or weighted difference strategy. A mixture model then can be described by combining Eqs. (4) and (5) as follows:

$$\mathrm{p}_{jk} = 1/(1 + \exp(-(b(w_1(\mathrm{S}_{j1} - \mathrm{S}_{k1}) + w_2(\mathrm{S}_{j2} - \mathrm{S}_{k2})) + a))). \quad (6)$$

Insofar as the weighting parameters (w_1, w_2) and the bias parameter (a) are all significant, we can conclude that the resulting choice proportions represent a mixture of strategies within the sample, within the individual, or both. Within the framework of Eq. (6) we can also assume that the bias parameter a may determine the likelihood of sampling information. The more extreme the value of a, the less likely the judge will even initially sample the information on the less important dimension. Thus, the model of Eq. (6) presents two types of weighting parameters. First, the bias parameter should relate directly to the strategic sampling of dimensional information and hence correspond to frequency of initial access to information. Second, the information weighting parameters should relate to the sampling of information during the valuation process and reflect TPAQ and additional frequency of access.

In Experiment 2, the strategy was to manipulate weighting within individuals by changing the task focus just as in Experiment 1. We will infer the use of a weighting model of choice when choice proportions are dependent on magnitudes of dimensional differences (Eq. (4)). The need for including the response bias parameter of Eq. (6) to model choice proportions will indicate the use of a more lexicographic strategy in which weight might be reflected in choosing whether or not to examine information. We predicted greater processing of information on the task relevant dimension. However, these process differences may occur in different ways, depending on the choice strategy. Paralleling the results of Experiment 1, those decision makers who follow a weighted additive or difference strategy should show greater looking TPAQ on the dimension receiving the higher weight. On the other hand, those decision makers who use heuristic strategies may show no differences in the time per acquisition because they are not weighting a value on a dimension. Instead, looking time differences for these persons should arise from a tendency to ignore information on the irrelevant dimension more often.

Unlike the judgment task of Experiment 1, the choice task of Experiment 2 may engage decision makers in an initial orientation and screening process. This process may reflect an initial sampling of information and a decision to eliminate alternatives from further consideration. One line of evidence for the existence of this stage within our choice paradigm would be if TPAQ on initial looks showed a different pattern than TPAQ on additional looks. Furthermore, evidence for screening would be found if participants engaged in a greater amount of strategic sampling in which information on one alternative is skipped altogether.

Experiment 2 paralleled closely Experiment 1. The major changes from that experiment were as follows: (a) verbal and math scores were presented for pairs of individuals, (b) participants selected the person from a pair who would be more likely to succeed in the major rather than rate each person, and (c) number of trials was expanded to include 100 pairs for each major. Once again, information was hidden on the screen in boxes and was uncovered by moving the mouse pointer into the box.

Method

Participants and Design

Participants were 53 students from a southern university, who received course credit for their participation. The basic design consisted of a 2×100 factorial combination of rating task (predict success in engineering or English major) and pair (100 pairs of scores). The order in which the two tasks were presented was counterbalanced so that half of the participants made choices of potential English majors first and half made choices of potential engineering majors first. The dependent variables included (a) the applicant who was chosen, (b) the types of transitions made as participants searched through information, and (c) looking time measures derived from process tracing.

Materials and Apparatus

All instructions and stimuli were presented on microcomputers, and responses were collected via the keyboard and mouse. The five levels of math and verbal scores used were 350, 430, 510, 590, and 670. The 100 choice pairs were constructed by combining the five levels in a pairwise fashion, resulting in a set of 25 stimuli, each with a math and a verbal score. These 25 stimuli were then combined to form a set of 300 unique choice pairs. Of these 300 pairs, 200 were eliminated because one alternative dominated the other (i.e., the alternative had a higher score on one dimension and an equal or higher score on the other dimension). Thus, each block consisted of 100 nondominated choice pairs. For each pair, one alternative had a high verbal and low math score (HVLM) and the other had a low verbal and high math score (LVHM).

Procedure

Instructions to participants paralleled those used in Experiment 1, except that participants were told to choose which of the applicants was more likely to succeed in the designated major. On each trial, four covered boxes appeared on the screen in a 2×2 matrix, each box containing the math or verbal score for Person A or Person B. Participants were told that the two scores for a person would be hidden behind boxes on the screen and that they should open the boxes by moving the mouse pointer into the box. When the mouse pointer moved into the box, the score was exposed, and when the pointer left the box, the score was hidden again. The labels “Person A” and “Person B” were printed to the left and right of the appropriate pair of boxes. There was no labeling outside of the boxes to indicate which box contained verbal and which contained math scores; however, on a given trial both verbal scores were either in the left set or the right set of boxes. After searching through the information, participants moved the mouse pointer to the label “Person A” or “Person B” marked outside of the boxes and clicked the mouse button to register their choice.

Before each set of choices, participants were given a practice trial to get acquainted with using the mouse to open boxes and record ratings. The experimental trials were presented in two blocks of 100 trials. There was a one minute rest period between presentation of each set. Presentation of score pairs was randomized within each block, and which score was presented in the left or right box on the screen was also randomized, so that the participants did not know prior to opening a box which box contained the math score and which contained the verbal score. This was done to encourage participants to examine all the information. Before making choices in the engineering major, participants were told that the major consisted of many courses involving highly complex math. Before making choices in the English major, participants were told that the major consisted of many courses involving reading difficult material.

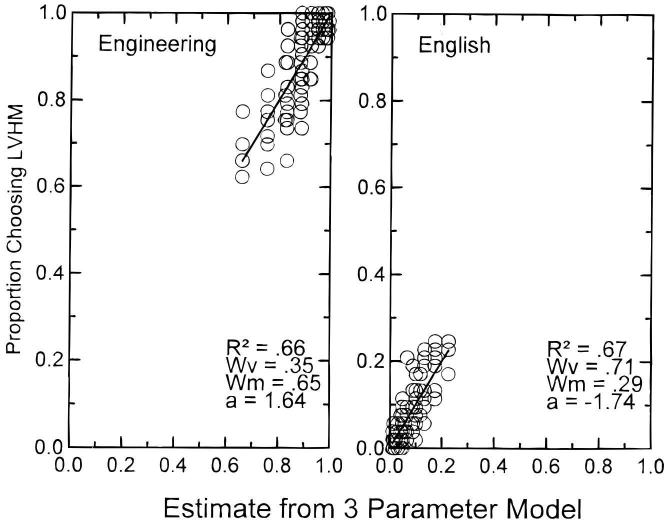

FIG. 5. Fit of three-parameter model to mean choice proportions for Experiment 2 (w_V = weight of verbal scores, w_M = weight of math scores, a = bias parameter). Weights and bias parameter values strongly shift across major.

Results

Choice Data

The manipulation of major produced a very large difference in the proportion choosing the LVHM over the HVLM persons. In the engineering major, the proportion choosing LVHM persons was .894, but in the English major, the proportion choosing LVHM persons was .084. Following Eq. (6), this difference might be attributed to a change in the relative weighting of verbal and math score differences, a change in the bias parameter for choosing the individual with the higher math score, or both. To investigate these possibilities, Eq. (6) was fit to the mean choice proportions of the participants separately for the English and engineering major using a nonlinear regression technique with a least squares loss function and a constraint that weights summed to 1.0. In each analysis, the full model that included the relative weighting parameter and the bias parameter fit the data significantly better than nested models that included only one of these parameters.

Figure 5 shows how model predictions compare to actual proportions for each major. In each case, roughly two thirds of the variance in proportions was explained by the model fit. When the data of Fig. 5 are combined, the fit accounts for 98% of the variance. Consistent with a shift in weighting explanation, the inferred relative weighting of verbal score differences was greater for the English major (w_V = .71) than for the engineering major (w_V = .35). Consistent with a lexicographic strategy, the bias parameter was positive for the engineering major and negative for the English major task.

The inclusion of both weight and bias parameters for the full set of participants may reflect the use of both weighting-based and lexicographic strategies within the same individual or the use of one or the other of these strategies across individuals. In an effort to tease these apart, we fit each individual’s set of 200 responses using a backwards stepping linear regression procedure. The full model included task focus (coded to reflect either the English or engineering major condition), a score difference variable on dimension 1 that corresponded to the magnitude of the difference of scores on that dimension, a score difference variable on dimension 2, and interaction terms involving score difference variables and focus. Model fitting proceeded by excluding the variable that would minimize change in R^2 at each step. Backward stepping stopped when exclusion of any variable resulted in a significant change in R^2 ($p < .05$).

Five patterns of models were of interest. The least interesting was the weight constancy model (WC), in which participants weight score differences but do not shift these weights across focus condition. Participants were classified as WC if the only terms included in the model were score difference variables. No participant fell into this category in Experiment 2.

A second model of interest is the weight shift (WS) model in which participants change the relative weighting of score differences across tasks. Participants were classified as WS if their regression model included only score difference variables and Score Difference × Focus interaction terms. Only four participants in Experiment 2 were classified as WS.

A third model corresponded to the bias shift (BS) model. This model is characterized by inclusion of only the focus term. For the most part, these participants simply chose the alternative with the higher score on the more important dimension. There were 16 individuals classified as BS in Experiment 2.

A fourth model corresponded to a bias shift with no shift in score differences. Participants were classified into the bias-shift-weight-constancy (BSWC) model if their regression equation included focus and at least one score difference variable, but did not include an interaction of focus and score difference. There were 9 such participants in Experiment 2.

Finally, participants could have responded to the shift in task focus by shifting both bias and weight parameters. Participants were classified as bias-shift-weight-shift (BSWS) if their equation included both the focus variable and a Focus × Score Difference interaction term. This was the modal category with 24 participants. Thus, although we did find a variety of separable strategies (as inferred from response patterns), the modal decision maker appeared to use a mixture of both shifting a global bias parameter (e.g., a lexicographic strategy) and shifting the relative weights of score differences.

Looking Time per Acquisition

The top four panels of Fig. 6 present the looking TPAQ results, with the top row corresponding to initial TPAQ and the second row corresponding to additional TPAQ. A four-way repeated measures ANOVA was conducted on the TPAQ for initial and additional looks. Because 6 participants had missing data in a cell of the design due to a consistent failure to look at a given piece of information, they were eliminated from the analyses in this section.

The significant main effect of stage, $F(1, 46) = 7.9$, $p < .01$, reflected the longer TPAQ for initial looks ($M = 560$) than for additional looks ($M = 516$). This difference may simply reflect the well documented decrease in time needed for rereading information (Hyona & Niemi, 1990). The only other significant main effect was an effect of dimension, $F(1, 46) = 8.8$, $p < .01$, which reflected slightly greater overall TPAQ for verbal scores

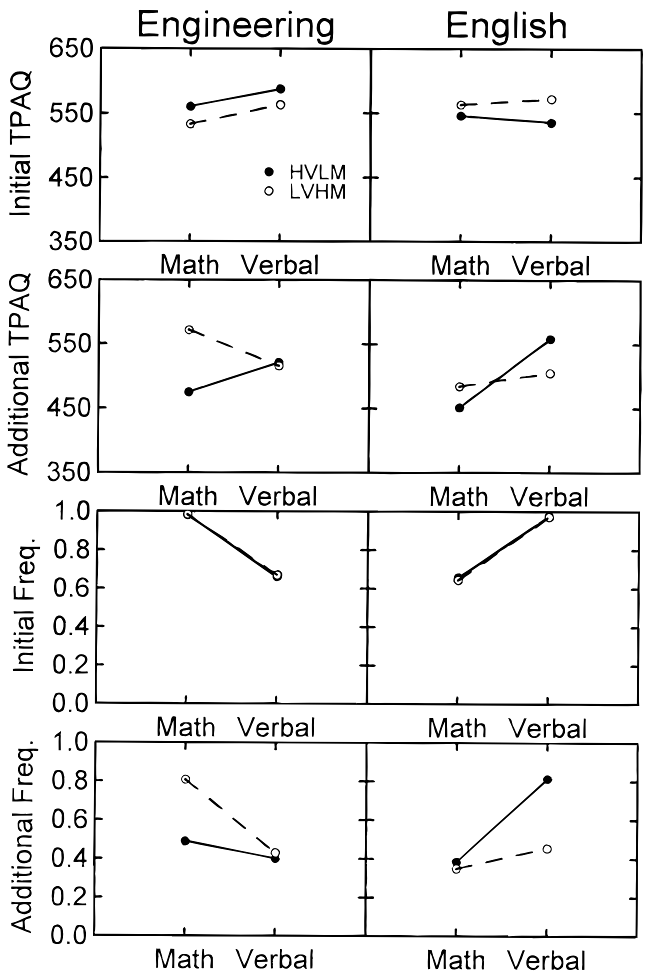

FIG. 6. Initial and additional time per acquisition (TPAQ) and frequency of access as a function of major, score, and person (Experiment 2).

($M = 550$) than for math scores ($M = 527$). This difference is consistent with the slightly greater overall weight inferred for verbal scores combining across the two tasks (see Fig. 5).

The critical test of a connection between looking time and weight is once again to be found in the Major × Dimension interaction, where TPAQ should be greater for the scores that receive greater weight in each major. This interaction did not achieve statistical significance, but it was in the predicted direction. There was, however, a significant three-way Stage × Major × Dimension interaction, $F(1, 46) = 12.8$, $p < .001$. Separate analyses on initial and additional TPAQ revealed that the Major × Dimension interaction was significant only for additional looks $F(1, 46) = 11.5$, $p < .001$. Thus, this analysis revealed that increases in weight were accompanied by increases in TPAQ only for additional looks and not for the initial look.

The analysis also revealed a significant Stage × Major × Person interaction, $F(1, 46) = 17.0$, $p < .001$. The two-way Major × Person interaction was significant for both initial and additional looks, but the interaction patterns were in the opposite direction. For initial looks, TPAQ was greater for the person who had the poorer score on the attribute most relevant to the major being assessed, $F(1, 52) = 18.6$, $p < .001$. Thus, the HVLM person had a higher initial TPAQ than the LVHM person when choosing between engineering majors (i.e., the solid line is above the dashed line in the top left panel of Fig. 6), but the reverse was true when choosing between English majors. For additional looks, TPAQ was greater for the person who had the higher score on the dimension most relevant to the major being assessed, $F(1, 46) = 5.9$, $p < .05$.

Finally, both the Person × Dimension and Stage × Person × Dimension interactions were significant for TPAQ, $F(1, 46) = 7.9$, $p < .01$, and $F(1, 46) = 24.4$, $p < .001$, respectively. Separate ANOVAs revealed that the Person × Dimension interaction was not significant for initial TPAQ but was significant for additional TPAQ. The interaction pattern for additional TPAQ reflected greater TPAQ for the higher score of each person, i.e. the math score for the LVHM and the verbal score for the HVLM persons.

Another way to think about the three combined two-way interactions for additional TPAQ is that they reflect greater TPAQ given the more relevant dimension to the more appropriate person within each task focus condition. Thus, the math score of the LVHM person gets the most additional TPAQ when choosing between engineering majors and the verbal score of the HVLM person gets the most additional TPAQ when choosing between English majors. This may reflect a verification process in which people spend more additional time on the information that leads to their choice, i.e., the more relevant score on the more appropriate person.

Frequency of Accessing Information

The bottom four panels of Fig. 6 present the mean frequency of initial and additional looks at each piece of information segregated by task focus. For initial looks, the highest mean possible for a piece of information is 1.0, which reflects always looking at that corresponding information on all 100 trials. The additional number of looks reflects the mean number of times a piece of information was examined following an initial examination. Thus, if the information was not examined initially, it did not contribute to the additional frequency statistic.

A four-way repeated measures ANOVA was conducted on the number of initial and additional looks. The only significant main effect was an effect of stage, which corresponded to greater initial frequency of access, $F(1, 52) = 69.9$, $p < .001$. The Major × Dimension interaction was significant and was in the predicted direction, $F(1, 52) = 293.6$, $p < .001$. Participants looked more often at information that was most relevant to the task. Thus, they looked more often at verbal scores in choosing between English majors and they looked more often at math scores in choosing between engineering majors. The interaction pattern was the same for both initial and additional looks, as reflected in a lack of a three-way interaction with stage. The lexicographic nature of the choices can be seen in the fact that information that was more relevant to the major was skipped only about 2% of the time, whereas information that was less relevant to the major was skipped about 34% of the time.

Two three-way interactions with stage were found, a Major × Person × Stage interaction, $F(1, 52) = 177.9$, $p < .001$, and a Person × Dimension × Stage interaction, $F(1, 52) = 143.3$, $p < .001$. As can be seen in the third row of Fig. 6, initial frequencies for the two types of persons were nearly identical for each task focus. Thus, these two large three-way interactions are mainly the result of large emergent interactions for additional frequencies. For additional frequencies, the Major × Person interaction was highly significant, $F(1, 52) = 206.7$, $p < .001$, and likewise the Person × Dimension interaction was highly significant, $F(1, 52) = 134.3$, $p < .001$. Like the TPAQ analysis, the three combined two-way interactions for additional frequencies indicated a clear pattern in which people tended to reaccess information on the more relevant dimension of the more appropriate person within each task focus condition. Thus, the LVHM person's math score was reexamined an average of .81 times in the engineering major while all other information was reexamined an average of only .44 times. A similar pattern occurred for reexamination of the HVLM person's verbal score in the English major as compared to reexamination of the other three scores (.81 vs .40).

Pattern of Looks

The looking time data described above is based on a set of participants who may have differed dramatically in their decision strategies. One major difference may have been the degree to which the individuals approached the task using a dimensionwise or an alternativewise strategy. To examine these differences, a PATTERN statistic was constructed by subtracting the number of dimensionwise comparisons from the number of alternativewise comparisons and dividing by the total number of comparisons (Payne *et al.*, 1988). This statistic can vary from −1.0, indicating the use of only dimensionwise comparisons to 1.0, indicating the use of only alternativewise comparisons. Overall, the group tended to be more dimensionwise in their processing of the information, with the PATTERN statistic significantly less than 0.0 ($M = -.267$). If we classify participants as alternativewise if PATTERN $>$ .10, dimensionwise if PATTERN $<$ −.10, and balanced if PATTERN falls between these values, then 44 of the 53 participants were classified as dimensionwise. The relatively large number of participants employing a dimensionwise strategy is consistent with the high usage of noncompensatory (bias shift) strategies (Payne *et al.*, 1988).

Discussion

Analyses performed on the choice data provided evidence for a mixture model (Eq. (6)), in which differences in task focus led to changes in the weighting of score differences as well as changes in a response bias parameter. Analyses of individual response patterns supported the conclusion that the mixture model characterizes the modal strategy employed by individuals (24 of 53 participants). Thus, participants showed a strong bias to choose the person with the higher score on the most relevant dimension, but they also tended to take into account the magnitude of score differences.

Unlike in the judgment task of Experiment 1, the results of Experiment 2 did not reflect greater processing times for the more relevant dimension. Furthermore, these initial TPAQs were much shorter in choice (546 ms) than in judgment (863 ms). These results are consistent with the idea that the initial TPAQ in choice is predominantly a orientation-screening stage rather than an evaluation stage (Russo & Leclerc, 1994). Further evidence for the screening aspect of initial TPAQ in choice was the significant Major × Person interaction. This interaction reflected greater initial processing times for those individuals whose scores were less appropriate to the task focus, i.e., the low math person for the engineering major. The greater initial processing of these individuals may have reflected a tendency to eliminate them from further consideration during an initial screening phase. Thus, the pattern of results for the initial TPAQ in choice did not parallel those for the initial TPAQ in judgment.

Like the additional TPAQ measure in the judgment task, additional TPAQ in choice reflected a Major × Dimension interaction, with greater TPAQ spent on the more relevant dimension. Thus, there is some evidence for the continual sampling model operating on additional looks in the choice experiment. This is consistent with additional looks falling within an evaluative stage of processing (Russo & Leclerc, 1994). However, the interpretation of the evaluation processes that occur during these additional looks must take into account the three combined two way interactions, which appear to point to greater processing of the more appropriate individual on the more relevant dimension. These results make sense within the mixture model. The greater initial screening time for the less appropriate person may have led to a tendency to minimally process this individual on additional fixations. Thus, the greater processing of the more important dimension only occurs for the appropriate person on additional acquisitions.

An alternative interpretation of this effect may be linked to a verification stage of processing, although in a different way than described by Russo and Leclerc (1994). Their experiment included many more alternatives so that the verification stage corresponded to a cursory examination of previously unexamined alternatives after a tentative decision had presumably been made. In our experiment, there were two alternatives and participants may simply have wanted to verify the good attribute of the individual they were choosing. This process seems similar to the bolstering process described by Srull and Wyer (1989) in the person impression literature. Bolstering is said to occur when one notices inconsistencies in the information describing a person. One way to resolve these inconsistencies is to go back and spend more time on the information that is deemed most relevant. In this sense, bolstering may be considered a weighting process that operates at a later stage of judgment. This weighting process may be tied to the idea that we are accountable for our choices and thus seek to have ready justification of our choice (Simonson, 1989; Tetlock & Boettger, 1989). Such a process could occur late in an evaluative stage or within a verification stage.

One of the clearest results from the process tracing data is that initial access is much more likely for the task relevant dimension. This Major × Dimension interaction on initial frequency supports the strategic sampling model and is consistent with the use of noncompensatory strategies in which a decision can be reached without examining all the information. The Major × Dimension interaction was equally strong for additional frequency of access, supporting the discrete sampling model. Thus, not only did people spend more TPAQ on the task relevant information, they went back to it much more often. Additional frequency of access also reflected the confluence of three two-interactions that reflected greater additional access to the more appropriate individual on the more relevant dimension. As with TPAQ, this pattern may reflect the latter stages of evaluation or a process within the verification stage.

The general picture that emerges from Experiment 2 is that choice engages more stages of processing than judgment. The initial stage may be used to gather relevant information and screen out inappropriate alternatives. The initial stage appears to lead to strategic sampling in which information on the less relevant dimension is unlikely to be sampled at all. In the evaluative stage, people go back more often and spend more time on the more relevant dimension of the more appropriate person. Such a pattern of additional frequency and TPAQ is consistent with both discrete and continuous sampling models of weight. It is unclear to what extent this latter sampling process is part of a verification rather than evaluative stage.

EXPERIMENT 3: EXAMINING LOOKING BEHAVIOR FOR DIFFERENT CHOICE STRATEGIES

Although Experiment 2 clearly demonstrated a strong shift in dimensional weights, the manipulation may have been so large as to prompt decision makers to become unidimensionally focused or lexicographic. This conclusion was supported by the large number of individuals who skipped information on the irrelevant dimension and the large number whose choices were insensitive to the magnitude of score differences.

The goal of Experiment 3 was to induce a larger number of participants to look at the information more fully so that separate analyses of those using different strategies could be developed. To this end, the majors were made less extreme. The engineering major was changed to an economics major and the English major was changed to a sociology major. By reducing differences in major, we hoped to sample a larger variety of strategies, both compensatory and noncompensatory. Our focus in analyzing the data was to classify individuals into different decision strategy groups primarily on the basis of their response patterns and then examine how process tracing measures differed across these groups.

Several basic relationships were predicted. First, we predicted that because of the greater processing demands of compensatory strategies, those participants who weighted score differences would have larger TPAQs and a greater number of acquisitions than those who did not. Similarly, those participants who followed an alternativewise processing pattern were predicted to have larger TPAQs and a greater number of acquisitions than those who followed a dimensionwise pattern. Furthermore, we reasoned that if the bias parameter of Eq. (6) corresponds to strategic sampling, then those participants whose results indicated the need for the bias parameter should show a greater Major $\times$ Dimension interaction on initial frequency of looks (i.e., they will skip information on the less relevant dimension more often). Based on the results of Experiment 2, we predicted that the initial TPAQ would reflect a screening and orientation stage, with similar initial TPAQs across dimensions within each major, but with more initial TPAQ spent on the less appropriate choice alternative. Finally, we expected that the Major $\times$ Dimension interactions found to reflect shifts in weight in Experiment 2 for additional TPAQ and additional frequency of acquisition would be found in Experiment 3. However, it was not clear whether this pattern would generalize across participants classified into different strategy groups.

Method

Participants were 120 students from a southern university, who received course credit for their participation. The method was identical to that of Experiment 2 except in the description of the majors. In Experiment 3, an economics major was substituted for the engineering major and a sociology major was substituted for the English major. Instructions for choosing between economics majors emphasized that economics requires extensive math background and skills to understand quantitative theories and analyses, but that it also requires good verbal comprehension. Instructions for choosing between sociology majors emphasized that sociology requires extensive verbal background and skills used to read difficult works, analyze them, and write papers, but that it also requires good math skills to analyze and understand data.

Results

Choice Data

Figure 7 displays a comparison of predictions from the mixed model of Eq. (6) to choice proportions for each major. In each condition, about one half of the variance of choice proportions was explained by the model, which was somewhat reduced as compared to the results of Experiment 2. This reduction of fit may have been due to the fewer number of extreme choice proportions in Experiment 3. When the data of Fig. 7 are combined, the fit accounts for 84% of the variance. The fit of Eq. (6) to the data revealed a shift in the inferred weighting parameters, with the weight of the verbal score being greater in the sociology major ($W_V = .60$) than in the Economics major ($W_V = .41$). As expected, these differences were not as extreme as those found in Experiment 2. The fit of Eq. (6) also included a significant shift in the bias parameter, which was positive for the economics major and negative for the sociology

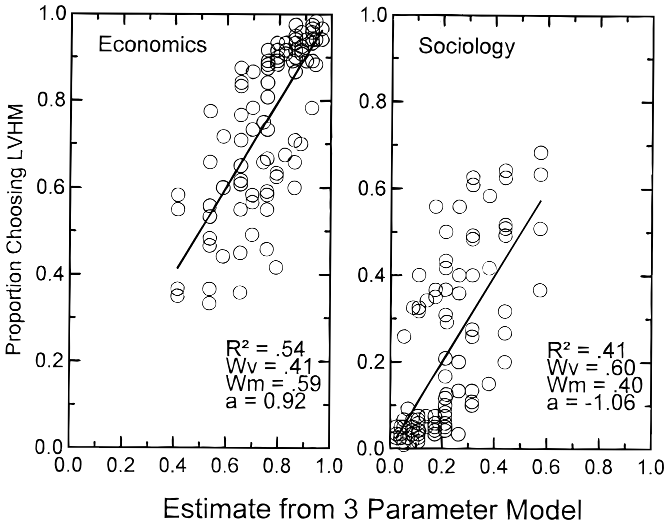

FIG. 7. Fit of three parameter model to mean choice proportions for Experiment 3 (w_V = weight of verbal scores, w_M = weight of math scores, a = bias parameter). Weights and bias parameter values shift across major.

major. Again as expected, the shift in the bias parameter was less extreme than in Experiment 2, possibly reflecting a reduced usage of a lexicographic strategy.

Looking Measures for Full Sample

In this section we describe how looking measures for the full sample in Experiment 3 compare to those for the full sample of Experiment 2. To facilitate this comparison, Table 1 shows the pattern of significance for analyses conducted on frequency of access and TPAQ

TABLE 1

Comparison of Full Sample Effects of Experiments 2 and 3

Effect	Experiment 2		Experiment 3	
	Freq	TPAQ	Freq	TPAQ
Stage	***	**	***	NS
Major	NS	NS	NS	NS
Person	NS	NS	NS	NS
Dimension	NS	**	NS	NS
Stage × Major	NS	NS	NS	NS
Stage × Person	NS	NS	NS	NS
Stage × Dimension	NS	NS	NS	NS
Major × Person	***	NS	***	NS
Major × Dimension	***	NS	***	NS
Person × Dimension	***	**	***	*
Stage × Major × Person	***	***	***	***
Stage × Major × Dimension	NS	**	NS	***
Stage × Person × Dimension	***	***	***	***
Major × Person × Dimension	NS	NS	NS	NS

Note. TPAQ, time per acquisition; time measure is in milliseconds. NS, effect or interaction is not significant.
*Effect or interaction is significant at .05 level.
**Effect or interaction is significant at .01 level.
***Effect or interaction is significant at .001 level.

TABLE 2

Classification of Strategy and Pattern of Processing

Pattern	Inferred strategy					
	WC	WS	BSWC	BSWS	BS	Total
Alternativewise	4	15	18	6	1	44
Balanced	0	4	15	16	1	36
Dimensionwise	0	3	11	15	11	40
Total	4	22	44	37	13	120

in both experiments. A cursory examination reveals that the pattern of significance was nearly identical for the two experiments. The corresponding patterns of main effects and interaction effects were also nearly identical. Below we briefly describe similarities and differences between the full sample results of the two experiments.

First it should be noted that the results for the key Major × Dimension interaction for frequency of access replicated across Experiments 2 and 3. This interaction reflected both the greater initial and additional accessing of task relevant scores. As in Experiment 2, there was a significant Stage × Major × Dimension interaction for TPAQ. This again reflected the emergence of the predicted Major × Dimension interaction for additional TPAQ.

The Major × Person and Stage × Major × Person interaction effects of Experiment 2 were replicated. The basic two-way interactions reflected greater processing of information for the person most qualified for the major. The three-way interaction for TPAQ once again reflected a reversal of the interaction pattern for initial and additional processing. In initial processing, the person with the poorer score on the focal dimension received significantly greater looking TPAQ, but the opposite was true in additional looks.

The Person × Dimension and Stage × Person × Dimension interactions were also replicated. Here again the pattern was that initially more time was spent on a person's poorer score, but in additional looks, more time was spent on a person's better score. Finally, it is interesting to note that the main effect of stage was not significant for TPAQ in Experiment 3. Thus, the amount of time of a typical initial look did not differ from the amount of time of a typical additional look.

Classifying Participants into Groups

The PATTERN statistic, which reflected the relative amount of alternativewise versus dimensionwise processing, was calculated for each participant in Experiment 3. Whereas PATTERN was significantly negative in Experiment 2, corresponding to a higher degree of dimensionwise processing, it did not significantly differ from zero in Experiment 3 ($M = -0.030$). Participants were classified into three groups according to their PATTERN statistic, with dimensionwise having PATTERN $< -.10$, alternativewise having PATTERN $> .10$, and balanced falling between these values.

As in Experiment 2, we fit each individual's set of 200 responses using a backward stepping linear regression procedure. Once again we sought to classify individuals into one of five models of interest. These models differed in whether score differences were weighted, whether these weights changed with task focus, and whether a bias parameter was needed. These distinctions were dependent on whether the focus term or the Focus × Score Difference terms were included in the model. Table 2 shows the classification of participants into strategies as well as how these related to the PATTERN statistic.

By far the smallest group was the WC, or weight constancy group, who weighted score differences but did not shift weights with task focus. The WS, or weight shift group, was similar to the WC group in not needing to include the bias term in their models, but differed in that they shifted weight with task focus. As one might expect, both weighting groups followed a predominantly alternativewise pattern of processing the information.

The next two groups distinguished in Table 2 showed evidence of both a bias parameter and weighting of score differences in making their choices. They differed in whether the weights of score differences shifted with task focus. Individuals in the BSWC, or bias-shift-weight-constancy, group did not appear to shift weights with task focus, but those in the BSWS, or bias-shift-weight-shift, group did. These groups appeared to differ in their pattern of processing the information. The BSWC group tended to be somewhat alternativewise in processing the information and the BSWS group tended to be somewhat dimensionwise; however, a substantial number in both groups fell into the balanced PATTERN group.

Finally, the last group differed from the other four groups in that these individuals' models did not include any score difference terms. Thus, persons in this BS, or bias-shift, group showed no tendency to weight score differences. Not surprisingly, persons in this group were strongly dimensionwise in their processing of the information.

Looking Behavior across Groups

Because of the very low number of participants classified into the WC group, these data were excluded from further analyses. In order to examine similarities and differences in the process measures across groups, separate MANOVAs were conducted for initial TPAQ, additional TPAQ, initial frequency, and additional frequency measures. These results are reported below, with the significance level set at $p < .05$.

Initial TPAQ. Figure 8 shows the pattern of initial TPAQ across experimental conditions for each of the four groups. The most striking aspect of the data is the similarity of the pattern across the four groups. Persons across different strategies spent similar amounts of time on each initial acquisition and showed a similar pattern of effects. The major systematic pattern evident in the data was reflected in a significant Major × Person interaction, with the less appropriate person to the task receiving greater TPAQ than the more appropriate person. Thus, the HVLM person was looked at longer on initial acquisitions than the LVHM person in the economics major, and the reverse was true in the sociology

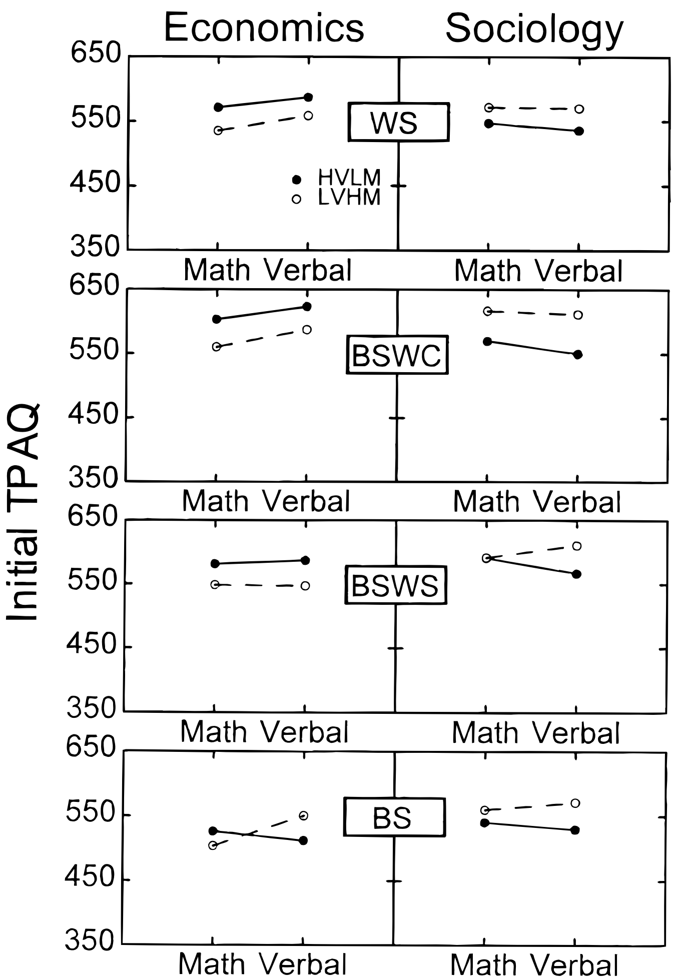

FIG. 8. Initial time per acquisition (TPAQ) as a function of major, score and person for participants classified into the weight-shift (WS), bias-sift-weight-constancy (BSWC), bias-shift-weight-shift (BSWS), and bias-shift (BS) strategies (Experiment 3).

major. These data are consistent with an initial screening-orientation stage as described by Russo and Leclerc (1994).

The repeated measures MANOVA revealed no main effect of group and only one significant interaction involving group, a Group × Major × Person interaction. Separate tests for each group found significant Major × Person interactions for all but the BS group. The lack of a Major × Person interaction for the BS group may be tied either to the failure of this group to weight score differences or due to the strong tendency of people in this group to skip information altogether, as described in the analyses below.

Additional TPAQ. Figure 9 shows the pattern of additional TPAQ across experimental conditions for each of the four groups. Unlike their initial TPAQ behavior, the groups clearly differed in their additional TPAQ behavior. A significant main effect of group indicated that the groups differed in their overall times

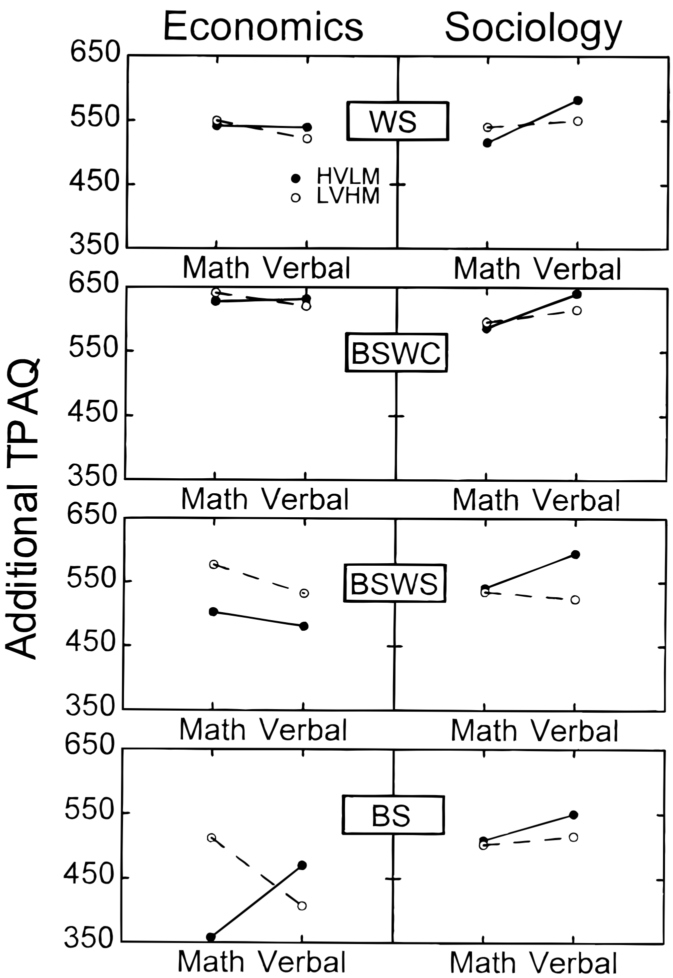

FIG. 9. Additional time per acquisition (TPAQ) as a function of major, score, and person for participants classified into the weight-shift (WS), bias-shift-weight-constancy (BSWC), bias-shift-weight-shift (BSWS), and bias-shift (BS) strategies (Experiment 3).

spent on these additional acquisitions. Additional TPAQ was greatest for the BSWC group (M = 617), was of intermediate value for WS (M = 542) and BSWS (M = 536) groups, and was least for the BS group (M = 478). This increased TPAQ may have reflected the greater processing requirements associated with weighting score differences and also with processing information by alternative.

Another salient difference across groups was reflected in a significant Group × Major × Person interaction. The WS and BSWC processors (top rows of panels of Fig. 9) showed very similar additional TPAQ for the two types of persons being considered, but the BSWS and BS processors spent more TPAQ on the person most appropriate to the major. The reduced processing time on the person less appropriate to the major is consistent with greater use of selective mechanisms for these two groups of participants.

Whereas initial TPAQ showed no tendency to look longer at the more relevant dimension, additional TPAQ did show this basic pattern of processing as reflected in a significant Major × Dimension interaction. There was, however, a significant Group × Major × Dimension × Person interaction. This interaction appears to be carried by a significant three way interaction for the BS group. In the economics major condition, the BS participants showed a strong Person × Dimension interaction that did not occur in the sociology major condition.

Initial frequency of access. Figure 10 shows the pattern of initial frequency of access across experimental conditions for each of the four groups. Not surprisingly, groups differed greatly on this measure. The most salient difference was reflected in the significant Group × Major × Dimension interaction. Although all groups showed the pattern of greater skipping of information on the less relevant dimension, this tendency was small for WS and BSWC participants, large for BSWS participants and extreme for the BS participants. These latter two groups then were very strategic in their sampling of information, often skipping over information on the less relevant dimension. The other major effect on initial frequencies was the significant main effect of group, which reflected the greater tendency to skip information for the latter two groups.

Additional frequency of access. Figure 11 shows the pattern of additional frequency of access across experimental conditions for each of the four groups. The significant main effect of group reflected the clearest difference among these groups. Information was reaccessed much more for the WS (M = .866) and BSWC (M = .913) groups than for the BSWS (M = .667) or BS (M = .536) groups. This difference may well reflect the

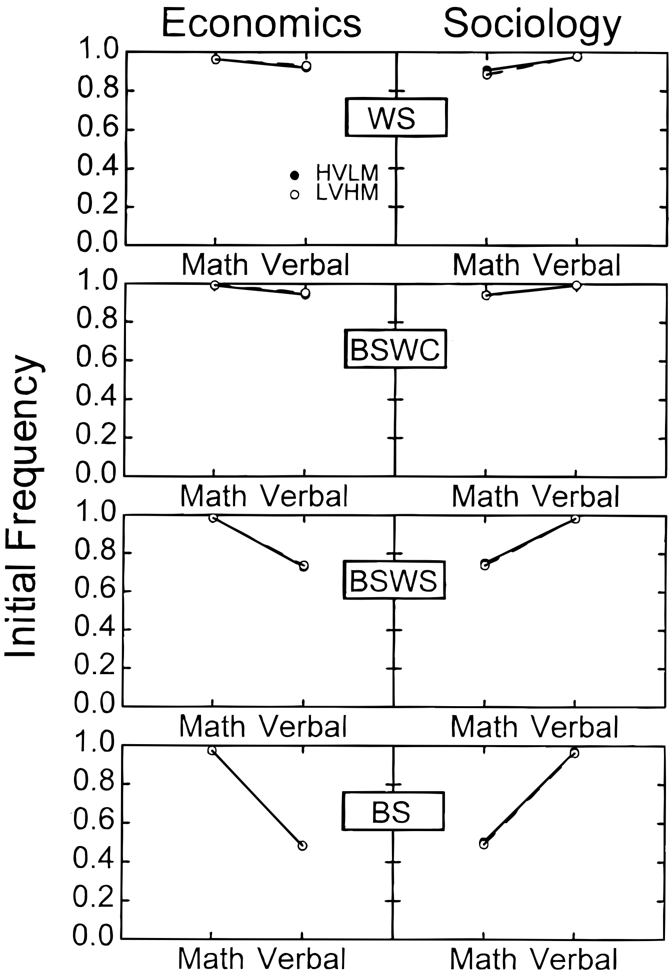

FIG. 10. Initial frequency of access as a function of major, score and person for participants classified into the weight-shift (WS), bias-shift-weight-constancy (BSWC), bias-shift-weight-shift (BSWS), and bias-shift (BS) strategies (Experiment 3).

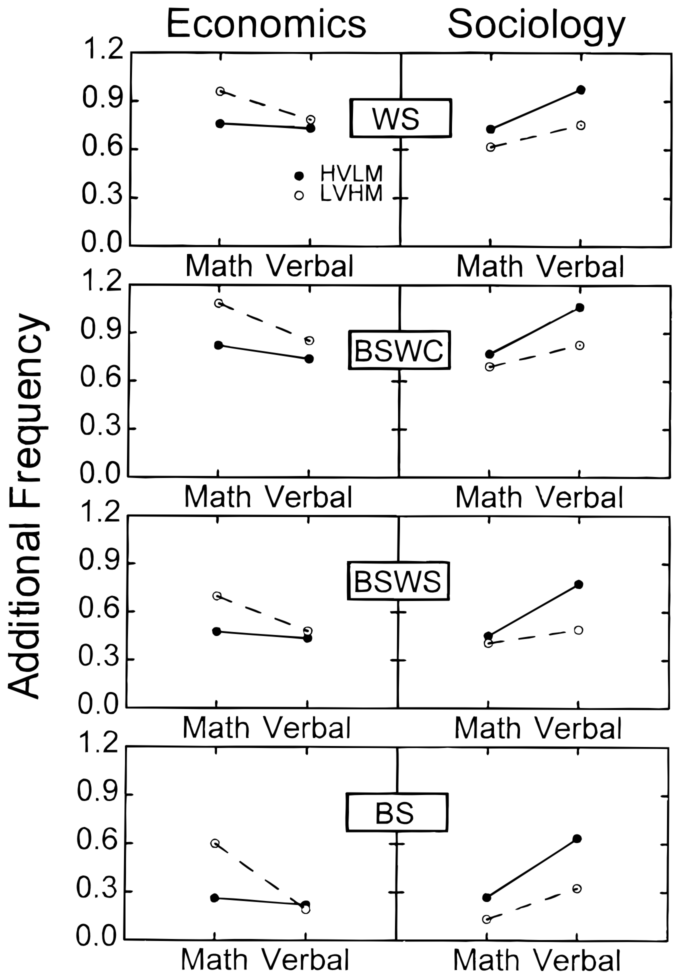

FIG. 11. Additional frequency of access as a function of major, score and person for participants classified into the weight-shift (WS), bias-shift-weight-constancy (BSWC), bias-shift-weight-shift (BSWS), and bias-shift (BS) strategies (Experiment 3).

greater tendency toward alternativewise processing for the former two groups. While groups clearly differed in the overall frequency of additional accesses, perhaps the most striking feature of these data is the similarity in the pattern of additional accessing across groups. As in Experiment 2, the greatest number of additional looks occurred on the relevant dimension of the person most appropriate to the major. The math score of the LVHM person was reaccessed with the greatest frequency in the economics major, and the verbal score of the HVLM person was reaccessed with the greatest frequency in the sociology major. This pattern was reflected in the combination of three two-way interactions (Major × Dimension, Major × Person, and Person × Dimension). Within our sampling framework, this pattern is consistent with the discrete sampling model. By going back to the more important dimension, this score is given greater weight.

A problem with this interpretation is that the same pattern holds for the BS participants, who do not appear to differentially weight score differences. The similarity of the interaction patterns across groups suggests that this effect may arise during a verification stage of processing, although verification processes differ somewhat from those described by Russo and Leclerc (1994). Having tentatively made a choice, the individual may wish to go back to the basis of that choice in order to verify that he or she is making the right decision. In this sense, the effect takes place post decisionally and therefore does not contribute to differential weighting of attributes. However, post hoc analyses show that the slopes of the functions in each panel of Fig. 11 for both HVLM and LVHM persons were significantly positive in the economics major and negative in the sociology major (with the only exceptions being the slopes for the HVLM person in the BSWS and BS groups). The greater additional access of the more relevant dimension for the less appropriate person is consistent with greater weighting of that dimension. Thus, this pattern may reflect both evaluation and verification processes.

Discussion

Experiment 3 closely replicated the pattern of results for Experiment 2 and thus demonstrated strong effects of task focus on both weighting and looking behavior. Because the manipulation of task focus was less extreme in Experiment 3, participants were more evenly divided across dimensionwise and alternativewise processing strategies, and patterns of responding tended to reflect the weighting of score differences for most participants. Thus, we were able to isolate four groups of participants whose choice patterns reflected different decision strategies. The groups showed some distinguishable differences in their patterning of looking behavior, but also showed some basic similarities.

One similarity was the use of an initial screening phase, evident in the greater initial time spent on the person more likely to be rejected and the failure to look longer at the more important dimension. Although three of the four groups showed this pattern in their initial TPAQ, they clearly differed in the degree to which they were willing to eliminate alternatives on the basis of partial information. The WS and BSWC groups tended to be much more compensatory in their behavior, rarely skipping information. The BS group was clearly the least compensatory, with the BSWS group in between.

Although shifts in the bias parameter tended to predict strategic sampling, it did so only for the BS and BSWS groups. The BSWC group appeared to follow a rather different strategy. First, these participants were much more alternativewise in their processing of information. Thus, the bias shift may have occurred late in the decision process and served the role of a tie breaker rather than occurring early in the process and guiding strategic sampling. The similarity in the pattern of the frequency of additional looks across groups is also noteworthy. Thus even though groups appeared to use very different bases for their decisions and different patterns of acquisition (e.g., WS versus BS groups), they showed a similar pattern of returning most often to the more relevant dimension on the individual more appropriate to the major.

GENERAL DISCUSSION

The overarching goal of the three experiments we conducted was to clarify the relationship between behavioral looking measures and the theoretical construct of weight. By manipulating task focus, we were able to induce participants to shift the weight they allocated to different dimensions. This allowed us to examine how behavioral looking measures change with shifts in weight. Our results demonstrated that changes in dimensional weight are accompanied by changes in looking behavior in both judgment and binary choice. These data provide strong support for the assertion that people look longer and more frequently at information to which they give greater weight.

Having established in these experiments convincing evidence that looking and weighting show clear correspondences, we would like to focus our general discussion on three issues. First, we will discuss evidence for at least two types of cognitive processes corresponding to the construct of weight. Second, we will discuss the time course of the weighting process. Finally, we will examine the basic link between looking behavior and decision processes.

Two Types of Weighting

Behaviorally, looking time on a piece of information can be broken down into constituent measures of the number of times that information is accessed and the time spent looking at the information on each acquisition. Both of these measures showed significant effects of the task focus manipulation. In some regards, these two measures might reflect the common underlying principle that weight corresponds to a sampling process so that the greater the sampling, the higher the weight. Information might be resampled by repeatedly accessing that information or it might be resampled within a given look. The former type of sampling would be reflected in the frequency measure and the latter in the TPAQ measure.

Although linking the constructs of weighting and sampling provides a coherent explanation for weighting effects on both frequency and TPAQ, the outcome and process measures suggest at least two ways in which weight is manifest in judgment and choice. According to the first, weight can be conceived as a modifier of stimulus values or stimulus differences, as expressed in the weighting parameters of Eqs. (1–6). In the judgment task, this might correspond to an anchoring and adjustment strategy in which the change in the adjustment of the current response is modified by weight, possibly through a repeated sampling process. This type of weighting is sensitive to differences in value. The good fit of the weighted additive model to the data of Experiment 1 provided validation for this conception of weight in judgment. In the judgment task of Experiment 1, shifts in weight were accompanied by shifts in both frequency and TPAQ.

The data of Experiments 2 and 3 were consistent with at least some participants in the choice situation following this type of weighting process in that their choice proportions were sensitive to differences in score magnitudes. The correspondence between increased dimensional weight and increased additional TPAQ and additional frequency of access provided evidence that in choice, as well as judgment, weight can be conceived as a modifier of stimulus value that is accompanied by increased looking behavior.

The choice data, however, provided strong evidence that many individuals follow a more qualitative strategy in making decisions. This type of process is not sensitive to score differences, but rather is captured in the bias parameter of Eq. (6). The shift in the bias parameter is consistent with several choice heuristics in which the role of dimensional weight might be to select a primary dimension upon which to make a decision. In strategies such as lexicographic choice (Tversky, 1969) or elimination by aspects (Tversky, 1972), dimensions are ordered in terms of importance, with the most important information being accessed and qualitatively evaluated first. In some cases, such as lexicographic choice, this qualitative evaluation of the information is totally insensitive to graded differences in magnitude. Because of their noncompensatory nature, these heuristic strategies may lead the decision maker to sample information only on the dimension that receives the greatest weight. The process tracing evidence supported the use of these types of heuristic strategies in a large number of participants, especially in Experiment 2. These participants were highly selective in their looking behavior, often failing to look even initially at information on the less important dimension.

In summary, weight may operate early in heuristic strategies as a selection mechanism or it may operate more directly on stimulus values or differences in values. The former application of weight is seen mostly in the differential accessing of information initially. The latter application of weight may operate both on TPAQ and on additional frequency of access. An exception to this general conclusion may be applied to the participants in Experiment 3 classified into the bias-shift-weight-constancy group. These participants were not selective in their initial frequency of access and looked long and often at the information. This type of late bias shift may reflect a tendency to break ties by choosing the alternative with the higher score on the most important dimension.

Time Course of Judgment and Choice

By breaking down the process tracing measures into initial and additional looks, we were able to examine the time course of the judgment and choice processes. In the judgment task, the Major × Dimension interaction was significant for all three initial measures. Thus, initial looks appear to include more than simply reading information into working memory, but also include valuing and weighting processes. Measures for additional looks in the judgment task were also sensitive to manipulation of weight in the same way, suggesting that weighting in judgment occurs early and continues through additional looks. In general, the process tracing data for judgment was consistent with use of only a single stage of processing, the evaluative stage within Russo & Leclerc's (1994) framework. This is not particularly surprising given that there was only a single alternative being presented on a trial.

The time course of the weighting process appeared to differ somewhat in choice and supported at least two stages of processing. The initial frequency and initial TPAQ data provided good evidence for an initial orienting or screening phase. First, because participants in the choice tasks were much more likely to follow noncompensatory strategies (especially in Experiment 2), frequency of initial access strongly reflected a screening phase so that information on the less relevant dimension was often skipped. Second, unlike those in the judgment task, initial TPAQ did not significantly reflect weighting differences: The looking times on initial acquisitions did not increase with increases in the importance of a dimension.

Differences between initial and additional looking behavior provide further support for a transition in processing from an orientation and screening stage to an evaluative stage (Russo & Leclerc, 1994). For example, the Stage × Major × Person and Stage × Person × Dimension interactions on TPAQ in Experiments 2 and 3 revealed a reversal of looking time patterns from initial to additional looks. Initially, participants tended to spend greater TPAQ on the negative dimension of a person; however, on additional looks they spent more TPAQ on the positive dimension of a person. Similarly, initially they tended to spend more TPAQ on the person least appropriate to the major, but this pattern was reversed on additional looks. The looking behavior on additional looks was similar to that on judgment trials and suggests an evaluative stage of processing. Thus, even with these very simple displays, choice appears to differ from judgment in that it includes an initial orientation and selection stage followed by an evaluative stage. One possible interpretation of this behavior is that at each point in the choice process, the decision maker is attempting to determine (a) what information to look at in the future and (b) whether to terminate information search and make a choice. Thus, the initial examination of negative information, especially on the most important dimension, might require additional time to determine whether the alternative should be accessed again or discarded (a screening function). The later focus on the positive attribute of the person more appropriate to the major could then reflect a weighting or a justification process in which decision makers are seeking to reaffirm their reasons for choosing this individual.

It is also possible that some of the latter looking behavior might have fallen into what Russo and Leclerc (1994) term a verification stage. In this stage, the individual has made a tentative choice and is searching alternatives with the purpose of verifying the correctness of this choice. In the Russo and Leclerc choice situation, there were many alternatives and the verification stage was defined primarily by the tendency to quickly search through the other alternatives in the set. Our two alternative situation did not lend itself to this type of behavior, but we did find a systematic pattern across all types of strategies to go back more often to the most important dimension on the person highest on this dimension. This behavior could be part of the evaluation stage, or it could be a type of verification in which one wishes to be assured of the reason for choosing this particular alternative.

Linkage between Looking Behavior and Decision Processes

Einhorn, Kleinmuntz and Kleinmuntz (1979) have pointed to the importance of combining outcome and process measures to develop a fuller picture of the decision process. Algebraic models of judgment and choice have proven to be good predictors of outcomes (Anderson, 1981), but they have not provided much in the way of process description. Process tracing studies, on the other hand, have provided a more detailed picture of the strategies involved in choice (Ford *et al.*, 1989), but they have often had to rely on an assumption that looking times reflect decision processes. The experiments reported here provide strong support for several links between looking measures and the conception of weight as both a selection mechanism and a modifier of value. The different patterns of processing that emerged for groups classified on their patterns of choices also helps to clarify some of the linkage between process and outcome models of choice. However, there are several reasons to caution against overgeneralizing these results.

Because outcome and process measures are both dependent variables, we have only demonstrated correspondences and not causal links between weight and looking behavior. We assume that it is weight that drives looking behavior rather than the other way around. Thus, we would not expect that forcing greater looking time on a piece of information would cause it to receive greater weight. Furthermore, even if we assume that increased weight may cause increased looking time, we have not demonstrated any necessity to this relationship. Thus, under different circumstances it is reasonable to assume that decision makers may weigh one piece of information while examining another piece of information. We examined a very simple experimental situation in which we could exercise a great deal of experimental control. Thus, these results do not necessarily extend to more complex types of information or alternatives that extend beyond two dimensions or choice situations that extend beyond two alternatives. On the other hand, the strong correspondences demonstrated here do provide some validity for the often assumed link between looking behavior and weight.

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology,* **102,** 543–561.

Böckenholt, U., Albert, D., Aschenbrenner, M., & Schmalhofer, F. (1991) The effects of attractiveness, dominance, and attribute differences on information acquisition in multiattribute binary choice. *Organizational Behavior and Human Decision Processes,* **49,** 258–281.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognitive approach to decision making in an uncertain environment. *Psychological Review,* **100,** 432–459.

Duffy, S. A. & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech,* **33,** 103–119.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review,* **86,** 465–485.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme information. *Journal of Personality and Social Psychology,* **38,** 889–906.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes,* **43,** 75–117.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review,* **94,** 236–254.

Hegarty, M. (1992). The mechanics of comprehension and the comprehension of mechanics. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 428–443). New York: Springer-Verlag.

Hyona, H., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica,* **73,** 259–280.

Lichtenstein, S., & Slovic, P. (1971). Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology,* **89,** 46–55.

Lichtenstein, S., & Slovic, P. (1973). Response-induced reversal of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology,* **101**, 16–20.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum.

Lopes, L. L. (1981). *Averaging rules and adjustment processes: The role of averaging in inference*. University of Wisconsin, Report 13.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, **92,** 383–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14,** 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, **43,** 87–131.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, **18,** 962–986.

Posner, M., & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. Rabbit & S. Dornic (Eds.), *Attention and performance V.* New York: Academic Press.

Rayner, K., & Morris, R. K. (1990). Do eye movements reflect higher order processes in reading? In R. Groner, G, d'Ydewalle & R. Parham (Eds.), *From eye to mind: Information acquisition in perception, search, and reading* (pp. 179–190). Amsterdam: North-Holland.

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9,** 676–696.

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, **21,** 275–290.

Russo, J. E., & Rosen, L. D. (1975). An eye-fixation analysis of multialternative choice. *Memory and Cognition*, **3,** 267–276.

Schkade, D. A., & Johnson, E. J. (1982). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, **44,** 203–231.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, **16,** 158–174.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of Explanations. *Psychological Bulletin*, **105,** 131–142.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review*, **73,** 596–605.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, **96,** 58–83.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, **23,** 86–112.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, **57,** 388–398.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, **76,** 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, **79,** 281–299.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, **95,** 371–384.

Wedell, D. H. (November, 1993). *Effects of different types of decoys on choice.* Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, D. C.

Received: July 26, 1996