Form and objective of the decision rule in absolute identification

J. D. BALAKRISHNAN
Purdue University, West Lafayette, Indiana

In several conditions of a line length identification experiment, the subjects' decision making strategies were systematically biased against the responses on the edges of the stimulus range. When the range and number of the stimuli were small, the bias caused the percentage of correct responses to be highest in the center and lowest on the extremes of the range. Two general classes of decision rules that would explain these results are considered. The first class assumes that subjects intend to adopt an optimal decision rule, but systematically misrepresent one or more parameters of the decision making context. The second class assumes that subjects use a different measure of performance than the one assumed by the experimenter: instead of maximizing the chances of a correct response, the subject attempts to minimize the expected size of the response error (a "fidelity criterion"). In a second experiment, extended experience and feedback did not diminish the bias effect, but explicitly penalizing all response errors equally, regardless of their size, did reduce or eliminate it in some subjects. Both results favor the fidelity criterion over the optimal rule.

In many formal descriptions of human perceptual performance, the perceptual judgment of a stimulus is divided into two logically distinct operations: encoding (or perceiving), and decision making (or response selection). Empirical support for these two stages is very strong, and few theorists have found any serious reasons to challenge this basic framework or to suggest an alternative one. At the same time, the amount of theoretical and empirical treatment of the two stages is far from equal. In many studies, the decision process is "accounted for" by estimating decision making parameters, but the ultimate purpose of this is merely to remove the contribution of decision processes from observable behavior, so that hypotheses about perceptual experience can be tested directly. The result is that although many theories seek to explain the relationship between physical stimuli and perception, few attempt to explain the relationship between perception and decision making.

When the research domain involves classification of some kind, a natural point of departure for discussions about decision making is the "optimal" decision rule—that is, the rule that maximizes the expected level of performance with respect to some quantitative performance measure. If percent correct is the targeted measure, the optimal decision rule becomes a "maximum likelihood" rule. On each trial, the decision maker computes the a posteriori stimulus likelihood,

$$P(S_i \mid \Psi) = \frac{f_{s_i}(\Psi)p_i}{\sum_{k=1}^{n} f_{s_k}(\Psi)p_k}, \qquad (1)$$

where f is the probability density function, Ψ is the perceptual effect (single or multivariate), and p_k is the prior probability of stimulus S_k. The response i is chosen that maximizes $P(S_i \mid \Psi)$. Because the optimal decision rule adds no free parameters to a model, it is sometimes the most convenient way to generate predictions about observable behavior from a model of perception. As a hypothesis in itself, however, this particular decision model may seem implausible, because it assumes that the subject has perfect knowledge of the probabilistic features (e.g., the noise model) of the perceptual system in the specific, and sometimes artificial, context of a laboratory experiment.

If the optimal decision rule is rejected on either theoretical or empirical grounds, the theorist is confronted with two new problems: first, to explain why the subjects do not perform at an optimal level, and second, to replace the optimal model with something else. In a recent study, Balakrishnan and Ratcliff (1996) reported evidence suggesting that subjects' decision rules in several kinds of two choice classification tasks are not based on stimulus likelihoods, but may be optimal nevertheless. Instead of the optimal decision rule, the subjects appeared to rely on a simpler, heuristic rule (a "distance-from-criterion" rule). Under certain conditions, this rule can perfectly mimic the optimal decision rule with respect to performance level. In this article, I report evidence that when there are more than two choices (i.e., absolute identification), the decision rule has a conspicuous and regular form and is decidedly suboptimal. When the number and range of the

stimuli are sufficiently small, this bias causes the percentage of correct responses to be highest in the center of the range (the "dominant center effect"). After outlining some of the effects of stimulus number and range on the size and pattern of response biases, I consider some possible interpretations of this result.

EXPERIMENT 1

In three between-subjects conditions, the stimuli to be identified were horizontal lines of different length. In two of the three conditions, the total range of the stimuli was relatively small, and the number of stimuli was either small ($n = 5$) or large ($n = 10$). In the third condition, the number was small ($n = 5$) and the spacing was large. Results from experiments in which both the number and range of the stimuli were large have been reported elsewhere, and will be considered later.

Method

Subjects. Thirty undergraduate students at Purdue University participated in partial fulfillment of an introductory psychology course requirement. Ten subjects performed in each of three between-subjects conditions. The subjects were prescreened for normal or corrected-to-normal vision, and the sessions lasted about 50 min.

Apparatus and Stimuli. The stimuli were displayed on a 14-in. CRT monitor at an approximate viewing distance of 40 cm. Horizontal lines differing in length and 1 mm in width were presented in the approximate center of the screen. In the small-range ($n = 5$) condition, the smallest line in the set was 8.6 cm and the spacing constant was 1.5 mm (total range = 0.6 cm). The smallest line in the 10-stimulus condition was 8.5 cm, and the spacing constant was 0.8 mm (total range = 0.7 cm). In the large-range condition, the smallest line was 1.7 cm, and the spacing constant was 1.0 cm. The subjects entered their responses using integer-labeled keys on the upper row of the computer keyboard.

Procedure. At the beginning of the session, the stimuli were presented together with their integer response labels, in order from shortest to longest. This demonstration was repeated as many times as the subject wished. Apart from this, subjects received no feedback during the experiment. After the stimulus demonstration was completed, each trial began with the presentation of a small, multicolored rectangle (3 × 6 mm), which served as an initial fixation point. This was followed after a 1-sec delay by the presentation of the stimulus object randomly chosen from the stimulus set, 1.8 cm below the fixation point and remaining illuminated for 1 sec. The horizontal center of the line was a uniform random sample within 2.5 cm to the left and right of the screen's center. The subjects were asked to try to determine which stimulus was presented—no special instructions or feedback were given about different types of error responses or their significance, and no pressure was imposed on the response time or the total number of trials completed by the end of a session. Breaks were allowed at any time following the end of a trial. On the average, the subjects completed 158 trials per stimulus in the two 5-stimulus conditions, and 49 trials per stimulus in the 10-stimulus condition.

Results and Discussion

Percent correct by stimulus size in the three conditions, combined across subjects, is shown in Figure 1. Of the three curves, the most consistent result was the inverted-U shape of the function in the small-range ($n = 5$) condition; this pattern was exhibited by each of the 10 subjects. Since the function is highest toward the center of the

range, this result will be referred to below as the *dominant center effect*. Percent correct in the other two conditions was less regular, with most but not all subjects showing a result similar to the one in the figure—a U-shaped or decreasing curve in the large-range condition, and generally but not always higher values in the center of the range in the small-range ($n = 10$) condition (cf. Murdock, 1960).

In order to examine the subjects' decision making strategies, the parameters of Thurstone's identification model ("Law of Categorical Judgment," Case V, Thurstone, 1927; Torgerson, 1958) were estimated.[1] The estimates were obtained with an optimization routine (gradient descent) to minimize sum squared error (SSE), and convergence to the same location was verified for several different initial parameter values. The resulting SSE values were 0.007 for the small-range ($n = 5$) condition, 0.014 for the small-range ($n = 10$) condition, and 0.009 for the large-range
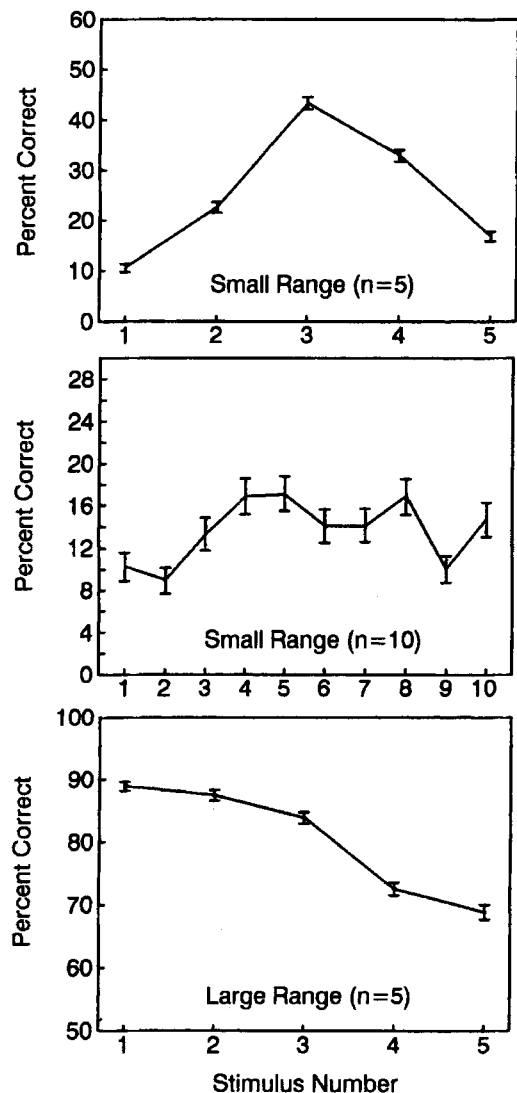

Figure 1. Percent correct by stimulus and condition in Experiment 1. Stimulus number is the integer label assigned to the stimulus. Physical sizes are given in the text.

condition. The average absolute deviation between observed and predicted values of the response proportions was less than .02, indicating a fairly close fit.

The spacing of the sensory distributions and the decision criteria are illustrated in Figure 2. Since the stimuli were equally likely, an optimal decision maker would place the criteria at the midpoints between adjacent distributions. In all three conditions, the pattern of criteria placement is qualitatively similar and clearly suboptimal: they are shifted outward, away from the midpoints between two distributions, toward the edges of the perceived length dimension (one exception occurs in the large-range condition for the criteria separating Responses 2 and 3). The overall effect of these response biases is to increase the frequency of responses toward the center of the range. At least to this extent, the effect is comparable to the "central tendency of judgment" noted previously in certain magnitude estimation and reproduction tasks (Hollingworth, 1913; Johnson, 1952; Laming, 1992; Woodworth & Schosberg, 1954).

To illustrate the substantial effects of the biases on the subjects' performance, the correct response percentages corresponding to the optimal decision rule for the three estimated distribution models are presented in the upper panel of Figure 3. In each case, the difference between the accuracy level implied by the optimal decision rule and the observed level is largest at the two extremes of the range, indicating a bias against these responses. In the two small-range conditions, the spacing between the sensory distributions is small, and the bias causes accuracy to be considerably higher with respect to the optimal rule in the center of the range. In the large-range condition, the same pattern of bias exists, but the criteria are much closer to their optimal locations, causing the shape of the percent correct function to approach that of the optimal decision rule. In the middle panel of Figure 3, the observed response proportions are compared with the proportions predicted by the optimal decision rule. Notice that in all three conditions, the biases increase the frequency of cen-
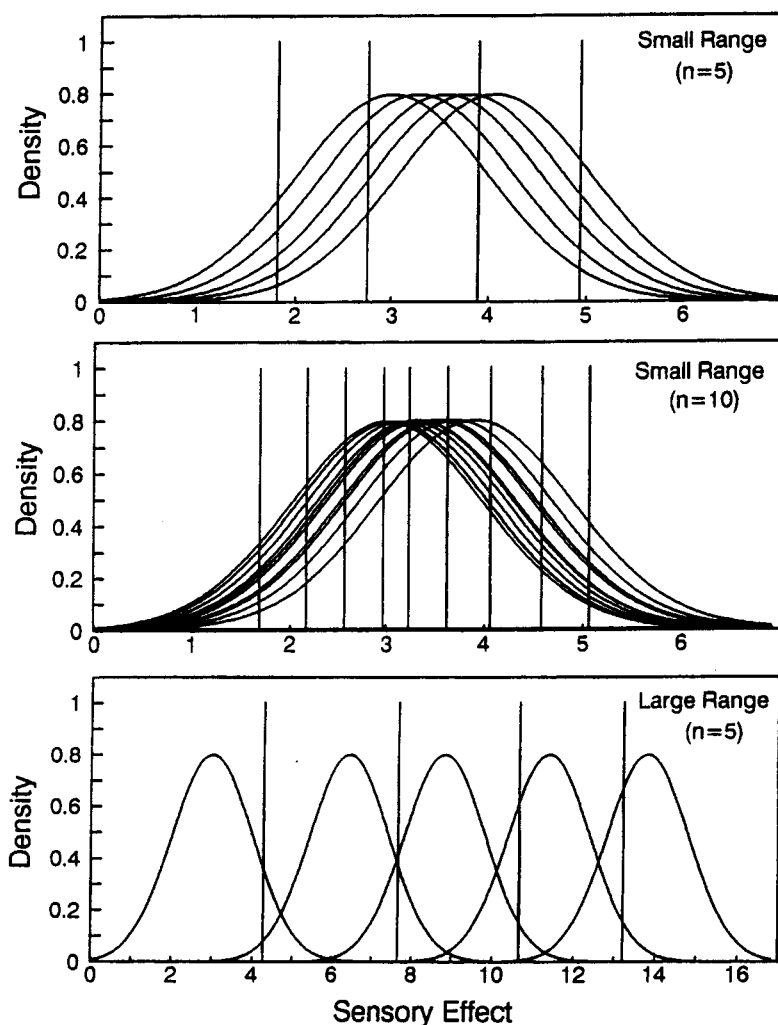

Figure 2. Criteria placements relative to the sensory distributions in three conditions of Experiment 1.
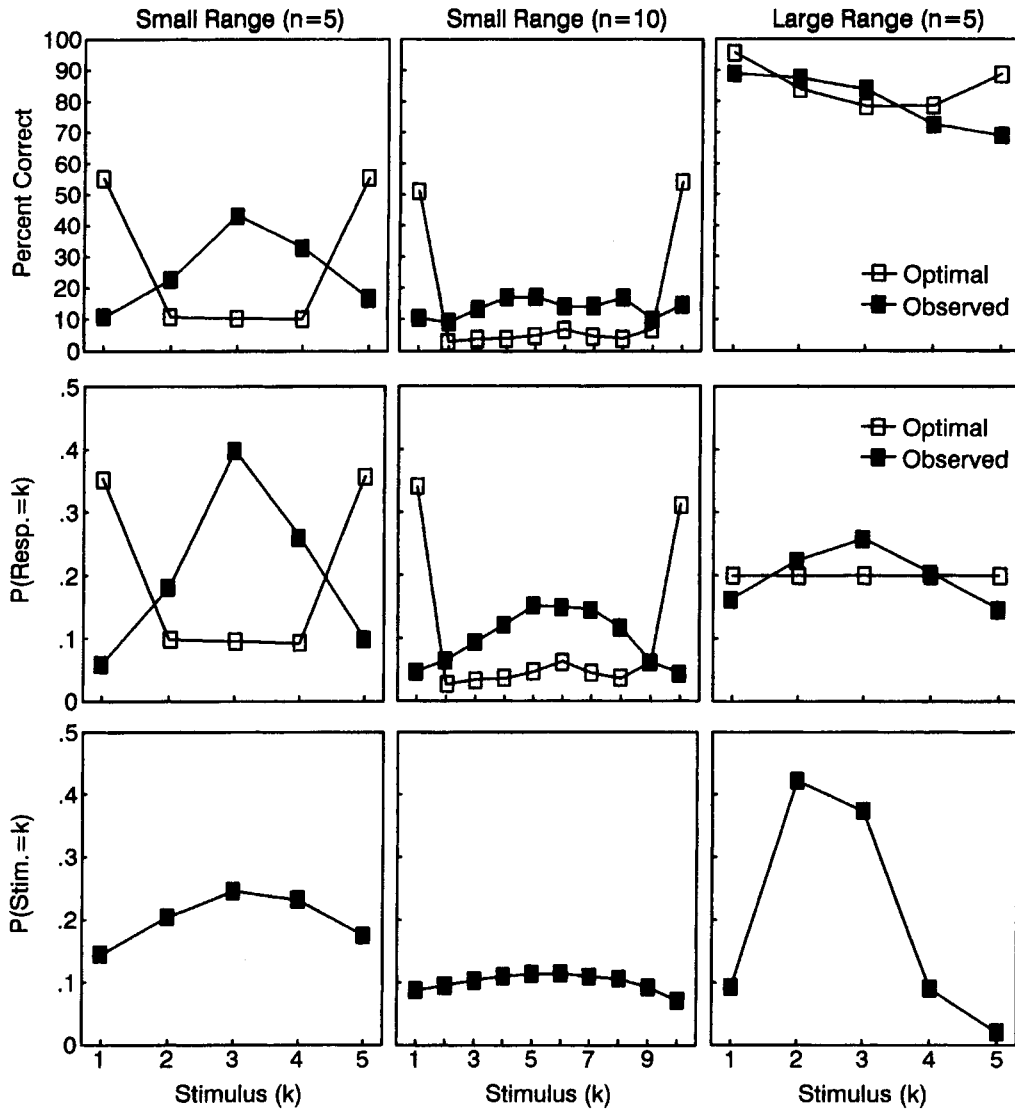
Figure 3. Upper panel: percent correct by stimulus when an optimal decision rule is applied to the distribution models of Figure 2. Middle panel: observed proportions of each response and proportions generated by the optimal decision rule. Bottom panel: a priori stimulus probabilities that would cause the observed criteria placements in Figure 2 to be optimal.

ter responses with respect to the optimal decision rule. In the large-range condition, an optimal rule would have produced virtually equal proportions of each response, and in the two small-range conditions, the bias causes the observed function to invert, with the center responses substantially more frequent than the edge responses.

The small size of the discrepancy between optimal and observed criteria placements in the large-range condition may be one reason why the bias against the edge responses has not been reported previously. In their comprehensive series of identification experiments, for example, Braida, Durlach, and colleagues explicitly considered the nature of the decision rule only in a large-range, 10-stimulus condition (Lippmann, Braida, & Durlach, 1976). Under these conditions there was no obvious pattern of deviation from

optimality. Inferring from this that the biases are necessarily weaker in any absolute sense when the discriminability of the stimuli is higher, however, would not be justified. In the bottom panel of Figure 3, the size of the bias is shown for each condition in Experiment 1, using another, equally valid measure—that is, the a priori stimulus probabilities that would have been necessary to cause the observed criteria settings to be optimal. In this "evidential" sense, the subjects' response biases appear to be considerably stronger in the large-range condition than in either of the two small-range conditions.

Interpretation of the Biases

Although there are several ways to define and compare the magnitude of bias in the different conditions of Ex-

periment 1, any measure derived from the Thurstonian representation of the data would indicate a significant departure from optimality and a systematic pattern of bias against the edges of the stimulus range. With this in mind, the next question is obvious: Why should the subjects adopt a decision making strategy that has a definite pattern, and at the same time a pronounced, detrimental effect on their performance level? I will consider three possible answers to this question. The first two assume that the subjects attempt to maximize their chances of a correct response, but misinterpret the actual parameters of the decision making context. I refer to these kinds of rules as *maximum likelihood criteria*. The third hypothesis assumes that the decision rule is designed to maximize a different performance measure altogether.

Misperceived stimulus probabilities. One way to explain the biases against the edges of the stimulus range was implicitly stated earlier: the subjects might systematically misinterpret the relative frequencies of the stimuli. In this case, the optimal model of Equation 1 can easily be revised so that, instead of being fixed by the experimenter, the prior probability values, p_k, in Equation 1 are free parameters. To account for the differences in the size of the biases between conditions, whatever effect causes the misperception in likelihood must depend strongly on the spacing of the stimuli. Ward and Lockhead (1971) and Treisman (1985) have each suggested mechanisms to account for other performance effects in identification tasks (i.e., sequential effects on response probability; see, e.g., Luce, Nosofsky, Green, & Smith, 1982; Mori, 1989; Ward & Lockhead, 1971) that might also reproduce the appropriate pattern of bias in perceived stimulus likelihoods. The main idea is that subjects mistakenly assume that the stimuli on previous trials are in some way informative about the stimulus probabilities on the current trial. In Ward and Lockhead's (1971) proposal, presenting a stimulus from one of the edges of the judgment dimension increases the perceived likelihood that a less extreme stimulus will be presented on the subsequent trial. Since the presentation of center stimuli does not increase the perceived likelihoods of extreme stimuli, it follows that the center stimuli are perceived to be relatively more likely overall.

Misperceived sensory distributions. Instead of misinterpreting the stimulus probabilities, the subjects might misrepresent the sensory effects of the different stimuli. In order to predict the observed bias pattern (e.g., large displacements of the criteria on the edges and small displacements in the middle), the simplest interpretation would be that the perceived locations of the sensory distributions are more closely spaced in the center of the range than they are at the edges. It is not obvious why this kind of misperception should occur, but the hypothesis turns out to be a special case of a model recently proposed by Lee and Ashby (1996).

Misconstrued objectives. Suppose that, consciously or otherwise, subjects assume that being close to the cor-

rect response when they do make an error has some intrinsic value. In other words, a small error between stimulus and response (an approximately correct response) is perceived as more desirable than a large error. Because the risk associated with any response increases with its distance from the center of the range, attempting to minimize the expected difference between the presented and reported stimulus would predict the observed response bias pattern. Larger biases in the large-range condition (see Figure 3) would be attributed to the larger errors associated with these stimulus conditions.

In classical communication theory, a decision rule that favors good approximation over perfect accuracy is referred to as a *fidelity criterion*, fidelity being the inverse of the discrepancy between received and transmitted signals (Shannon, 1960). The fidelity criterion is implemented by defining a loss function in which the size of the penalty increases with the size of the error. In an identification task with n stimuli, the subject chooses the response j that minimizes

$$\Omega_j = \sum_{k=1}^{n} \omega_{j,k} P(S_k \mid \Psi), \qquad (2)$$

where $\omega_{j,k}$ is the subjective cost of responding j when the stimulus is k, and the other variables are defined as in Equation 1 above. If the $\omega_{j,k}$ are constant, the resulting model is identical to Equation 1; more generally, these values are a nondecreasing function of some suitable measure of the difference between each stimulus pair i, j.

Since the difference between these three hypotheses is the interpretation of an accepted model fit (Thurstone's model), and not a comparison of different models, it is difficult to rule out any of them on the basis of a single study. In the remainder of this article, I will consider empirical results that are either directly predicted by, or more consistent with, the third hypothesis.

Monotonicity of the Subjective Loss Function

Once the perceived stimulus frequencies or the perceived distribution means are allowed to be free parameters in Equation 1, almost any observed deviation from optimality could be reproduced perfectly by this model.[2] The fidelity criterion, on the other hand, makes a stronger, empirically testable prediction about the identification data. Let i and j denote two stimuli in an n-stimulus identification task. If subjects adopt the fidelity criterion, then for each j, the cost values $\omega_{j,k}$ that produce the closest fit to the data should be nondecreasing in k for all $k > j$.

Since the stimuli in the experiment were equally spaced, the model can be restricted further by assuming that there are n unique cost values, one for each unique physical difference value in the identification set. Even without limiting the number of cost parameters, it should be noted that these parameters control only the decision rule, and hence the fit of Thurstone's model is always an upper bound on the fit of any fidelity criterion model. Adding
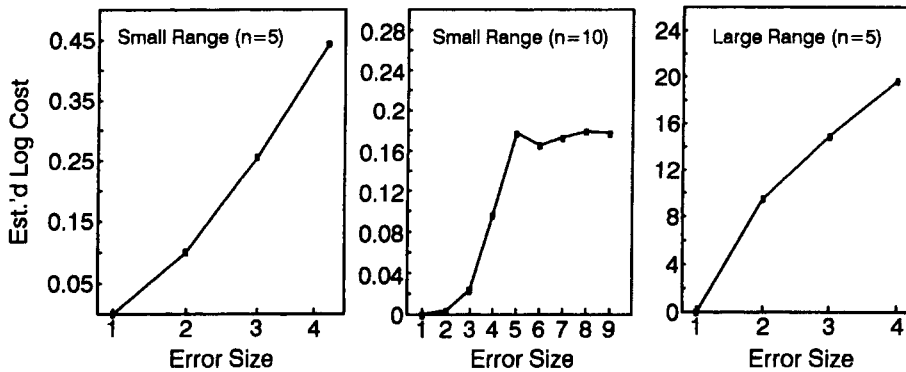
Figure 4. Estimates of the subjective costs of each response error, log(ω_i), for the three conditions of Experiment 1.

this "symmetry" constraint, however, substantially reduces the complexity of the parameter estimation problem. Finally, whatever assumptions are made about the cost values, it is easy to see that for any $\alpha > 0$,

$$\min_j \left[\sum_{k=1}^{n} (\alpha \omega_{j,k} + \beta) P(S_k \mid \Psi) \right]$$

$$= \alpha \min_j \left[\sum_{k=1}^{n} \omega_{j,k} \, P(S_k \mid \Psi) \right]$$

$$+ \beta \sum_{k=1}^{n} P(S_k \mid \Psi), \tag{3}$$

for all j and all distribution models $P(S_k \mid \Psi)$. Hence, the set of estimated $\omega_{j,k}$ values are nonidentifiable up to any linear transformation of the set. Without consequence, then, $\omega_{j-k=0}$ can be set to zero and $\omega_{|j-k|=1}$ to 1, and the remaining $n-2$ cost parameters can be estimated, together with the $n-1$ distribution means in Thurstone's Case V model. Except for estimation error, the estimated subjective costs should be strictly increasing with the error size that they represent.

Results of this analysis are shown in Figure 4 for the three conditions of Experiment 1. The estimates were obtained with the use of a least squares criterion. Instead of the raw estimates of the ω_k parameters, the figure shows their natural logarithms, since this makes their slopes easier to compare. *SSE* values were 0.011 for the small-range ($n = 5$) condition, 0.019 for the small-range ($n = 10$) condition, and 0.045 for the large-range ($n = 5$) condition. In the two 5-stimulus conditions, the ω_k estimates are sharply increasing; in the 10-stimulus condition, the function is not strictly increasing, but the violations are very small and appear as the function reaches an apparent asymptote. Overall, the predictions of the model appear to be well satisfied by the data. The fact that the slope of the function is considerably higher in the large-range condition has no strong implications, but it is consistent with the assumption that the subjective costs also increase with the physical size of the error when errors are compared between conditions.

EXPERIMENT 2

In the single-session design of Experiment 1, the subjects had a relatively limited amount of experience with the task, and they were not given any feedback during the course of the experiment about the accuracy of their responses.[3] Previous research has shown that both practice and feedback increase the overall performance level of subjects (e.g., Braida & Durlach, 1972; Eriksen, 1958), and one reason for this may be an increase in the accuracy of the subjects' representation of the stimulus context. If the response biases in Experiment 1 were caused by a systematic pattern of errors in the subjects' representation of the stimuli, it seems reasonable to expect that feedback and experience would help to reduce or eliminate them. In the first condition of Experiment 2, paid subjects performed the same small-range identification task of Experiment 1 in several sessions and were given feedback at the end of each trial.

A second condition was designed to test a similar prediction of the fidelity criterion. Specifically, if the biases in Experiment 1 were caused by subjectively higher rewards for smaller errors, an external payoff system that explicitly rewards the subject for treating all errors equally should compete with this tendency, reducing or eliminating the bias. Subjects in the second condition performed the same identification task as that in Condition 1 in several sessions. In addition to the feedback, however, these subjects received a single positive payoff for all correct responses, and a single negative payoff for all incorrect responses. Since experience in one condition might affect performance in another condition, different subjects were recruited for the two conditions. Furthermore, to ensure that the objective, monetary incentives would not be higher in one condition than in the other, the subjects' payments in the payoff condition were not contingent on their total point scores.

Method

Subjects. Eight Purdue University students performed several 1-h sessions (five to seven sessions each) and were paid $5.00/h for their participation. All subjects reported normal or corrected to normal vision.

Apparatus and Stimuli. The stimulus displays were identical to those of the small-range, five-stimulus condition in Experiment 1.

Procedure. Four subjects participated in each of the two conditions. The procedures were the same as in Experiment 1, with the following exceptions. In both conditions, the response given and the correct responses were presented in separate, labeled rows immediately after the subjects entered their responses, and this information remained on the screen for 1 sec. In the payoff condition, the accuracy feedback was followed by a new screen showing the point score for the trial and the total points accumulated. The subjects received 2 points for a correct response and −1 point for any incorrect response. The subjects in the accuracy condition were given the same instructions as were the subjects in Experiment 1, which emphasized accuracy over speed; the subjects in the payoff condition were asked to attempt to achieve as high a score as possible. The subjects in both conditions were told that the stimuli were equally likely.

Results and Discussion

Percent correct by stimulus size, subject, and condition is shown in Figure 5. Overall, performance levels were higher in this experiment than in the small-range condition of Experiment 1 [39.3% correct overall, versus 25.4% in Experiment 1; $t_{diff}(30825) = 23.31$, $p < .001$], and the increase could not be attributed to changes in decision making strategies (predicted percent correct for the optimal decision rule in Experiment 1 was only
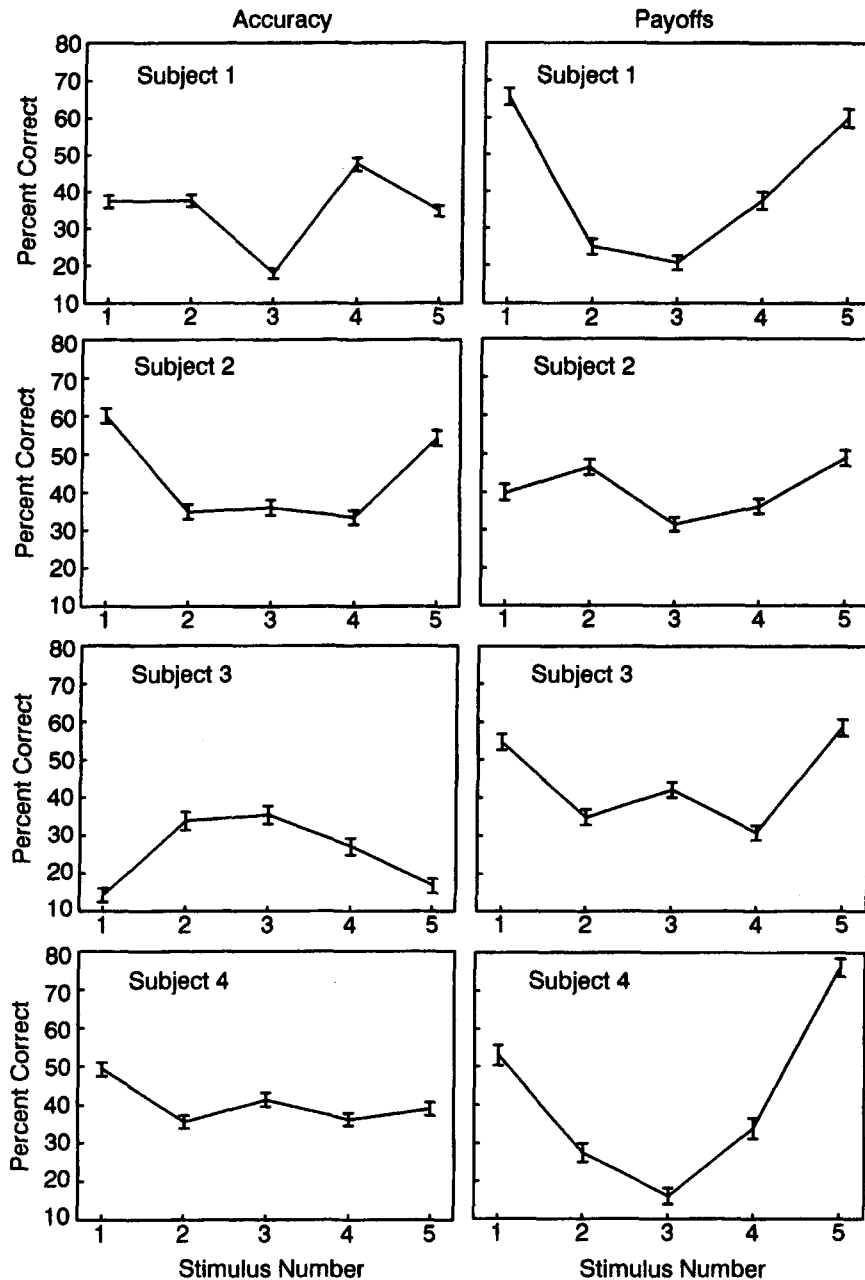

Figure 5. Percent correct by stimulus, subject, and condition in Experiment 2.

Table 1
Estimated Parameters of Thurstone's Case V Model for Each Subject in the Accuracy and Payoff Conditions

Subject	SSE	N	μ_1	$C_{1,2}$	$d_{c-c(opt)}$	μ_2	$C_{2,3}$	$d_{c-c(opt)}$	μ_3	$C_{3,4}$	$d_{c-c(opt)}$	μ_4	$C_{4,5}$	$d_{c-c(opt)}$	μ_5
						Accuracy Condition									
1	0.002	4,408	0.000	−0.325	−0.592	0.534	0.705	−0.086	1.048	1.107	−0.157	1.480	2.431	0.670	2.043
2	0.002	2,930	0.000	0.252	−0.243	0.789	1.138	−0.052	1.592	2.060	0.102	2.326	2.950	0.258	3.058
3	0.005	2,039	0.000	−1.029	−1.163	0.269	0.077	−0.410	0.707	1.060	0.336	0.742	2.039	1.123	1.090
4	0.001	4,102	0.000	−0.013	−0.326	0.625	0.892	−0.052	1.261	1.946	0.348	1.933	3.048	0.695	2.773
						Payoff Condition									
1	0.010	2,193	0.000	0.400	0.161	0.478	1.039	0.145	1.310	1.535	−0.109	1.978	2.398	0.075	2.668
2	0.004	3.048	0.000	−0.380	−0.690	0.619	0.925	−0.071	1.372	1.774	0.117	1.943	2.674	0.414	2.578
3	0.002	2,868	0.000	0.132	−0.242	0.747	1.050	−0.103	1.589	2.137	0.162	2.390	2.941	0.161	3.169
4	0.003	1,698	0.000	0.077	−0.256	0.665	0.819	−0.116	1.204	1.268	−0.371	2.075	2.232	−0.285	2.960

Note—$C_{i,i+1}$ = estimated criterion separating response region i from region $i+1$; $d_{c-c(opt)}$ = difference between the observed criterion $(C_{i,i+1})$ and the optimal criterion, $(\mu_i+\mu_{i+1})/2$; N = total sample size; SSE = least squares fit of the model.

28.4%). However, one subject (Subject 3 in the accuracy condition) performed at about the same level as did subjects in the previous experiment, and these data clearly exhibit the dominant center effect. Thus, in both experiments, the appearance of this effect depended on the overall performance level. The increase between the two experiments could have been due to any of several factors, including the different populations (paid vs. unpaid subjects) or the feedback (cf. Braida & Durlach, 1972; Eriksen, 1958). Since the general pattern of response biases observed in Experiment 1 did not depend on the overall accuracy or the presence or absence of the dominant center effect, the higher accuracy in Experiment 2 does not imply anything about the nature of the subjects' decision making strategies in this experiment. The subsequent analyses will therefore directly compare the size and pattern of the response biases between the two conditions.

Least squares estimates of the parameters of Thurstone's Case V model are given in Table 1. For convenience, the table also lists the signed differences between the estimated criteria and the optimal criteria implied by the estimated distribution model ($d_{C-C(opt)}$). Negative values for the first two numbers and positive values for the second two numbers in the rows of the table would indicate a consistent bias toward the center responses. Three of the 4 subjects in the accuracy condition show the complete pattern of increasing bias with distance from the center. The remaining subject (Subject 1) violates the pattern in only one case. In the payoff condition, 2 subjects (1 and 4) clearly are not biased toward the center responses, and 1 subject (3) shows this effect to a relatively small degree. A t test on the (unsigned) deviation between observed and optimal criterion locations (i.e., on the average of the $|d_{C-C(opt)}|$ estimates between conditions) was marginally significant [$t_{diff}(30) = 2.209, p = .051$]. Although the response biases were not completely eliminated in the payoff condition, then, they did not have the size or consistent pattern of the accuracy condition or the three conditions of Experiment 1. Note that any effect at all of the payoffs on the size or presence of the bias would be difficult to explain if subjects in both conditions were maximizing the perceived likelihood of a correct response.

To illustrate the effects of the payoffs on the size of the biases, the estimated percentage of trials on which the subjects made a suboptimal response (i.e., a different response was objectively more likely to be correct) is shown in the rightmost column of Table 2. In the other columns of the table, the direction of the biases is illustrated by subtracting the observed proportion of response $k + 1$ when response k was optimal ($\delta_{k,k+1}$) from the proportion of response k when response $k+1$ was optimal ($\delta_{k+1,k}$). (Note that one of these two values must be zero.) If the values in the first two columns ($k = 1, k = 2$) are negative and the values in the second two columns ($k = 3, k = 4$) are positive, the bias favors the center responses. Again, with one exception (Subject 2), the results suggest less deviation from the optimal decision rule in the payoff condition. The average of these (unsigned) bias estimates (i.e., the average of $|\delta_{k+1,k}|$) was .100 in the accuracy condition, and .048 in the payoff condition [$t_{diff}(30) = 2.296, p = .029$].

Finally, for comparison with Experiment 1, the estimated loss functions in the fidelity criterion decision model are shown in Figure 6, using once again the logarithm of the estimated costs in order to reduce the acceleration rate of the functions. SSE values were less than 0.05, and the average absolute deviation between observed and predicted values of the matrices was less than 0.03 across fits. Note that each of the four estimated functions is strictly increasing in the accuracy condition. In the payoff condition, the curves are not always monotonic, and in general their slopes are closer to their optimal value (0).

SUMMARY AND CONCLUSIONS

In several conditions of an absolute identification task, a surprisingly large proportion of the subjects' identification errors appeared to be due to a poor decision making strategy. Instead of choosing the response associated with the maximum a posteriori likelihood, subjects seemed to systematically favor responses in the center of the stimulus range over responses on the edges. The phenomenon was very robust: it was unmistakable in each of three between-subjects conditions (Experiment 1) and in each of four highly experienced subjects (in the accuracy condition of Experiment 2). When the number and the range of
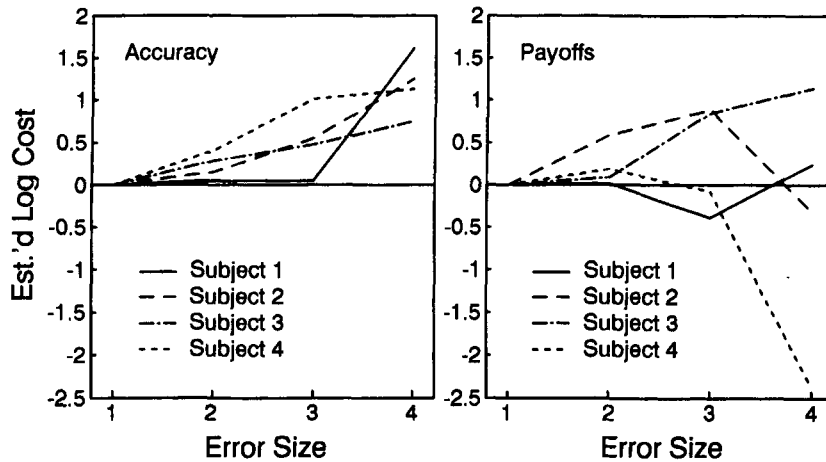
Figure 6. Comparison of the estimated subjective costs of the response error in the accuracy and payoff conditions of Experiment 2.

the stimuli were both relatively small, the effect of the bias was a higher percentage of correct responses in the center of the range than on either of the two edges. An optimal decision rule would have produced the exact opposite result.

Since the decision rule is clearly neither optimal nor arbitrary in form, the finding seems to offer a chance to understand in a more detailed way the operations and objectives of the decision making process. Two major classes of decision rules were defined that would not only fit the data (Thurstone's model does this already), but also explain the origins of the effect. The first class assumes that subjects base their responses on the maximum likelihood criterion of the optimal decision rule, but for one reason or another they systematically miscalculate it. The second class assumes that the subjects adopt a fundamentally different criterion, a fidelity criterion, which takes into account not only the accuracy of the response, but also the size of the response error.

The distinction is purely a matter of interpretation: once a necessary condition on the fidelity criterion is satisfied (monotonicity of the loss function), virtually any single data set could be explained by either model. The problem is therefore to find a suitable plausibility test. Along these lines, I have shown that the bias against extreme responses does not disappear with experience or feedback, both of which would be expected to improve the subjects' representation of the identification context and hence their ability to calculate likelihoods accurately. The bias was reduced or eliminated, however, by explicitly penalizing all errors equally, which does not change the optimal decision rule, but may draw attention to the fact that small errors are not more desirable from the experimenter's point of view. Neither result is unequivocal, but both favor the fidelity criterion over the maximum likelihood criterion.

Relationship to Other Identification Models

There are several other performance effects in the perceptual identification literature, and some of them have also been attributed to decision making. Before conclud-

ing, I comment briefly on how the present results might fit into the larger picture of identification developed in previous research.

One concept that plays a key role in several general theories of human identification performance is "criterial" or "memory noise" (e.g., Durlach & Braida, 1969). Broadly defined, this is any variation in the mapping between sensory effects and responses from one trial to the next. In one form or another, criterial noise has been held to account for a wide range of findings in identification, including sequential effects (Treisman, 1985), the "channel capacity" in unidimensional identification (Durlach & Braida, 1969; Marley & Cook, 1984), and the decreased discriminability of stimuli in the center of the range when the total range of the stimuli is large (the "bow" or "resolution-edge" effect, Durlach & Braida, 1969; Luce et al., 1982).

When the form and objective of the decision rule are explicitly stated, criterial noise may be attributed to changes in the subject's internal representation of the stimuli from trial to trial (Durlach & Braida, 1969), or in the perceived relative likelihoods of the stimuli (e.g., Treis-

Table 2
Estimates of the Proportion of Biased Responses
for Each Stimulus and Combined Across Stimuli

Subject	$\delta_{2,1}-\delta_{1,2}$	$\delta_{3,2}-\delta_{2,3}$	$\delta_{4,3}-\delta_{3,4}$	$\delta_{5,4}-\delta_{4,5}$	δ_T
	Accuracy Condition				
1	−.134	−.027	−.049	.149	.360
2	−.028	−.013	.025	.051	.116
3	−.276	−.146	.119	.287	.828
4	−.069	−.014	.089	.124	.295
	Payoff Condition				
1	.037	.038	−.028	.016	.119
2	−.129	−.019	.032	.087	.267
3	−.046	−.024	.038	.031	.139
4	−.055	−.030	−.095	−.059	.238

Note—$\delta_{i,j}$ = the proportion of trials on which the subject responds i when the optimal response is j; δ_T = the total proportion of suboptimal responses.

man, 1985; Ward & Lockhead, 1971). In the fidelity criterion model (Equation 2), the parameters of the functions used to compute $P(S_j \mid \Psi)$ would be changed from constants to random variables, whose expected values may or may not equal the parameters of the objective likelihood function. Assuming that the objective of the decision rule is to minimize the expected size of the error would not be incompatible with this kind of mechanism, since the computation of this expected value also requires a representation of the identification context and the ability to compute stimulus likelihood. In this respect, there is no reason to suspect any strong connection to exist between the biases observed in this study and other effects previously associated with the decision process. The implications for quantitative models of identification are therefore relatively straightforward: a comprehensive model of the decision process would not only be stochastic, but would also include an additional set of parameters that define the set of subjective penalties associated with the different response errors.

REFERENCES

BALAKRISHNAN, J. D., & RATCLIFF, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 615-633.
BRAIDA, L. D., & DURLACH, N. I. (1972). Intensity perception; II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, **51**, 483-502.
DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **53**, 1270-1287.
ERIKSEN, C. W. (1958). Effects of practice with or without correction on discriminative learning. *American Journal of Psychology*, **69**, 350-358.
HOLLINGWORTH, H. L. (1913). The central tendency of judgment. *Archives of Psychology*, **4**, 44-52.
JOHNSON, D. M. (1952). The central tendency of judgment as a regression phenomenon. *American Psychologist*, **7**, 281.
LAMING, D. (1992). Two categories of contextual variable in perception. *Behavioral & Brain Sciences*, **15**, 572-573.
LEE, W. W., & ASHBY, F. G. (1996). *Decision rules in stimulus identification*. Manuscript submitted for publication.
LIPPMANN, R. P., BRAIDA, L. D., & DURLACH, N. I. (1976). Intensity

perception: V. Effect of payoff matrix on absolute identification. *Journal of the Acoustical Society of America*, **59**, 129-134.
LUCE, R. D., NOSOFSKY, R. M., GREEN, D. M., & SMITH, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics*, **32**, 397-408.
MARLEY, A. A. J., & COOK, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical & Statistical Psychology*, **37**, 136-151.
MORI, S. (1989). A limited-capacity response process in absolute identification. *Perception & Psychophysics*, **46**, 167-173.
MURDOCK, B. B., JR. (1960). The distinctiveness of stimuli. *Psychological Review*, **67**, 16-31.
SHANNON, C. E. (1960). Coding theorems for a discrete source with a fidelity criterion. In R. E. Machol (Ed.), *Information and decision processes* (pp. 93-126). New York: McGraw-Hill.
THURSTONE, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 273-286.
TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
TREISMAN, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology*, **29**, 175-230.
WARD, L. M., & LOCKHEAD, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, **9**, 73-78.
WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology* (2nd ed.). New York: Holt.

NOTES

1. Estimates of parameters in Thurstone's unequal variance model (Case III) were also obtained with the same procedures. However, since the conclusions would be the same, only the results of the simpler model are reported here.
2. The necessary and sufficient condition is that a solution exists for the set of equations defined by $f(C_k \mid S_k) p_k = f(C_k \mid S_{k+1}) p_{k+1}$, where $k = 1, \ldots, n$ and C_k is the criterion separating response k and response $k+1$ in Thurstone's (1927) model.
3. All decision models must assume that the response is chosen prior to any feedback and hence must be based on some internal model of the identification context. The presence or absence of feedback therefore does not have any bearing on the issue of the choice of decision rule. Presumably, subjects in Experiment 1 began with some representation of the stimuli derived from the initial stimulus demonstration and did or did not adjust this during the course of the experimental session.