Dimensional Commensurability and Cue Utilization in Comparative Judgment

by

Paul Slovic

and

Douglas MacPhillamy

Oregon Research Institute
Eugene, Oregon

Research Bulletin
Vol. 11 No. 14

November 1971

Revised
November 1972

Dimensional Commensurability and Cue Utilization

in Comparative Judgment[1]

Paul Slovic[2]

and

Douglas MacPhillamy

Oregon Research Institute

Abstract

In a series of five experiments, subjects compared pairs of students with respect to potential college GPA. Both students had scores on one common dimension (e.g., English Skills) and one unique dimension (e.g., Quantitative Aptitude for Student A and Need to Achieve Success for Student B). The results indicated that dimensions were weighted more heavily in the comparison when they were common than when they were unique. Cautioning subjects not to increase the weight of the common dimension did not reduce the effect, nor did "correct answer" feedback with rewards for accuracy. In addition, the effect was substantial whether or not the common and unique dimensions had equal means and standard deviations. The results are congruent with a growing body of research that documents man's limitations as an information processor.

Dimensional Commensurability and Cue Utilization in Comparative Judgment

Paul Slovic

and

Douglas MacPhillamy

Oregon Research Institute

Most of the important stimuli about which judgments and choices are made are multidimensional. A multidimensional stimulus is characterized by two or more component attributes, dimensions, or cues. For example, tones can be characterized by pitch and loudness; hospital patients by temperature, pulse rate, and results of various lab tests; airplane landings by ceiling, visibility, and the amount of fuel remaining at touchdown; etc. A general problem of interest is that of understanding how a decision maker combines information of varying relevance from the several cue dimensions of a stimulus to arrive at a quantitative judgment of that stimulus or a choice among two or more stimuli.

Recent work has indicated that many types of judgments about multidimensional stimuli can be predicted by a simple linear combination of the individual cues (Slovic & Lichtenstein, 1971). That is, the process of weighting and combining cues can be approximated by an equation of the form,

$$J(X) = b_1X_1 + b_2X_2 + \ldots + b_kX_k + \varepsilon \quad , \qquad (1)$$

where J(X) is the judgment; X_i is the scale value of the i^{th} cue; b_i represents the weight or importance of the i^{th} cue, i = 1, ..., k; and ε is a global error term, indicative of the inherent unreliability in human judgment.

Our interest in the judgment process centers around the manner in which certain structural characteristics of the judgment task influence (a) the specific weights employed, and (b) the ability of the judge to weight cues according to his beliefs about their importance. Several kinds of structural properties have previously been shown to influence cue utilization. These include such factors as the order of presentation of the cues to the judge, the manner in which the judge is asked to express his response, cue format, and cue variability (Slovic & Lichtenstein, 1971). Even when these structural characteristics should be irrelevant to cue utilization, they seem to have strong effects on the ways that judges integrate information.

One particular concept that has been useful in explaining the effects of structural factors upon cue utilization is the notion of "cognitive strain" (Bruner, Goodnow, & Austin, 1956). It appears that some structural characteristics of cues add to the complexity of the information-processing task required of the judge and cause him to change his cue utilization systematically in an attempt to reduce the strain on memory, attention, and other components of reasoning.

The structural effect investigated in the present study is the degree to which commonality of cue dimensions influences cue utilization in situations requiring comparative judgments. Consider a task in which a judge must compare two students with respect to the criterion of potential college grade point average (GPA). He is given each student's score on two cue dimensions on which to base his judgments. One dimension is common to both students but the other dimension for each student is unique. For example, Student A may be described in terms of his scores on tests of Need for Achievement (NAch) and English Skills (Eng), while Student B may be described by his scores on Quantitative Ability (Quant) and Eng.

How might commonality affect cue utilization? We hypothesize that cue dimensions will have greater influence on comparative judgments when they are common to each alternative than when they are unique to a particular

Basic instructions. All subjects received the following general instructions about the task:

> This experiment is part of a study about the decision-making process. We are interested in how people use different kinds of information in arriving at decisions. Your task will be to use several pieces of information to arrive at some decisions in a realistic decision-making problem, in this case, making predictions about students' grades on the basis of test scores. The tests which you will use to make these predictions are:
>
> NAch (Need for Achievement)--This is the student's score on a questionnaire designed to determine his motivation to achieve success in school. Some of the questions on the test were:
>
> a. It is important for me to get good grades. (Yes or No)
> b. I feel that my future peace depends upon my accomplishment.
> c. I would rather be a good student than have an active social life.
> d. I enjoy relaxing only after completion of a job well done.
>
> The average score on this test is 50.
>
> Eng (English Skills)--This is a test of vocabulary skills and the ability to comprehend the English language. (Note: there are no foreign students in the sample you will be seeing.) The average score on this test is 100.
>
> Quant (Quantitative Aptitude)--This is a test of computational and arithmetic skills and the ability to use quantitative reasoning in solving problems. The average score on this test is 500.
>
> "In each case, some part of the test information will be missing. In no case will other relevant information (such as I.Q., health, personality, etc.) be available. While these limitations will make your task difficult, this is often the position we are placed in as real decision makers. We don't have all of the information we'd like, yet we must make a decision anyway. This is precisely what we will be asking you to do.
>
> "The whole experiment should take about one hour. But there is no time limit. Please work at a comfortable pace and try to make the best decision you can each time."

Stimuli. Each subject judged 15 practice pairs of stimuli and 90 test pairs. Within the 90 test pairs, there were 30 for which NAch was the common dimension (Set N), 30 with Eng in common (Set E), and 30 with Quant in common (Set Q). These three sets were randomly interspersed in the test booklets.

Within each set, the test scores were constructed to be normally distributed, with prespecified means and standard deviations.

Half of the subjects judged students for which the three test scores were measured in different units (i.e., different means and standard deviations). For the other subjects, all three cues had equal means (500) and equal standard deviations (150). Table 1 illustrates this facet of the design. Examples of stimulus pairs drawn from the equal and unequal-unit conditions are shown in Table 2.

Insert Tables 1 and 2 about here

Information about bias. In addition to varying the equality of units for the three cue dimensions, one other experimental manipulation was performed. One-half of the subjects were given the following information along with the other instructions:

"Each pair of students will have one test in common. People sometimes pay too much attention to that test simply because it's easier to compare the students on the test they have in common. Try not to pay more attention to a test simply because both students have a score on it. Instead, use each test according to your belief about its relative importance to the judgment you are making."

Subjects. The subjects were 120 paid volunteers from the University of Oregon. Thirty subjects were assigned to each of the four experimental conditions defined by crossing the two levels of cue measurement (equal vs. unequal units) with the two levels of information (informed about overweighting of common dimensions vs. uninformed). There were approximately equal numbers of men and women in each condition. Subjects within each condition were tested simultaneously in a group setting.

Data analysis. The data analysis was based upon the fact that linear models such as Equation 1 have been found to do a good job of predicting a judge's responses in multiple-cue tasks (Slovic & Lichtenstein, 1971). However, the present analysis differs from the usual application of linear models in two ways. First, the data being analyzed are not the subject's judgments of individual stimuli but rather his judgments of the difference between pairs of stimuli. Second, the typical linear model would estimate just one weight for each cue dimension, across the 90 pairs of stimuli. However, the common-dimension hypothesis predicts that the weights for a particular cue will not be invariant across the the three linear common dimension differs. Therefore, equations were computed for each subject--one equation for each stimulus set. These equations were:

$$\underline{\text{Set N:}}\quad d(AB) = n_1N_A - n_2N_B + q_3Q_A - e_4E_B \tag{2}$$

$$\underline{\text{Set E:}}\quad d(AB) = e_1E_A - e_2E_B + q_4Q_A - n_3N_B \tag{3}$$

$$\underline{\text{Set Q:}}\quad d(AB) = q_1Q_A - q_2Q_B + e_3E_A - n_4N_B \tag{4}$$

where d(AB) represents the judged difference between the GPAs of Students A and B (when A was chosen, the difference was coded as positive; choice of B received a negative sign); the terms n_i, e_i, and q_i are standardized regression weights; and N_A, N_B, E_A, E_B, Q_A, and Q_B are standard scores for Stimulus A and Stimulus B for dimensions NAch, Eng, and Quant, respectively. For purposes of this analysis, we define Student A in Set N as the student who has scores on NAch and Quant, while Student B has scores on NAch and Eng. In the test booklet, however, A and B were randomly interchanged.

The weights n_i, e_i, and q_i can be viewed as a measure of the relative importance of each of the four items of information (two common scores plus two unique scores) in determining the judged difference between students. To facilitate the interpretations of these weights in terms of relative importance, the stimuli were constructed so that the four items of information were orthogonal (uncorrelated) across the 30 pairs in each stimulus set.

The main hypothesis under study is that information will be weighted more highly when it is common to both students than when it is unique to one of the students. This hypothesis implies that the sum of the absolute values of the two weights for the common dimension in a set should exceed the sum of the weights for that same dimension in the two other stimulus sets where that dimension was unique.

Thus, we hypothesize that

$$|n_1| + |n_2| > |n_3| + |n_4|$$

$$|e_1| + |e_2| > |e_3| + |e_4|$$

and

$$|q_1| + |q_2| > |q_3| + |q_4| .$$

These sums of weights for common and unique dimensions served as the dependent variables for the primary data analyses.

The analyses just described are based on the subjects' judgments of the quantitative difference, d(AB), between pairs of stimuli. In addition, in order to evaluate the relative weighting of the common dimension in the basic binary choices, the judges' quantitative responses were recoded as 1 (Student A selected) or 0 (Student B selected) and the analyses outlined above were then repeated. This will be referred to as the "choice analysis."

Results

The validity of the linear models (Equations 2, 3, and 4), computed for each subject and each stimulus set, was evaluated by correlating each model's predictions with the subject's responses, across the 30 pairs of stimuli. These correlations averaged .857 in the analysis of quantitative comparisons and .779 in the choice analysis, sufficiently high to warrant further analysis of the standardized regression weights according to the design given above.

The upper half of Table 3 presents the mean weights for the common and unique dimensions in this experiment. The analysis based on the quantitative comparisons of the students is presented in the first two columns; the analysis of the binary choices is given in the third and fourth columns. Each entry in the table is the mean of sums of two weights. For example, the .89 value for Quant as a common dimension, given in the eighth row of Table 3, is the mean of $|q_1| + |q_2|$ and the corresponding value of .81 for Quant as a unique dimension is the mean of $|q_3| + |q_4|$.

Insert Table 3 about here

The statistical significance of these descriptive data was evaluated by means of a fixed-effects analysis of variance with two between-subject variables (information about bias and cue measurement) and two within-subject variables (common vs. unique dimensions, and type of common information--NAch, Eng, or Quant). A separate analysis was done for the weights derived from the binary-choice analysis.

The data in Table 3 confirm the hypothesis that common information is given greater weight than unique information. The main effect of the common vs. unique dimension was quite reliable statistically both for the quantitative comparison, $F(1,116) = 21.6$, $p < .001$, and for the binary choice, $F(1,116) = 13.2$, $p < .001$. Another index of the consistency of the effect for the quantitative comparison is the fact that about 68% of the subjects gave more weight to a dimension when it was common than when it was unique.

The effect of the common dimension on the binary choices was as great as the effect on the quantitative comparisons. This indicates that the common dimension exerts its special influence on the basic choice itself. Had the effect not been present in the analysis of the binary choices, this would have indicated that the influence of the common dimension occurred only during the quantification of the difference between students.

None of the two-way interactions between the common vs. unique dimension and the other experimental variables (information about bias, cue measurement, or type of common information) were statistically significant. In other words, the common-dimension effect was equally strong for each of the three types of common information, as can be seen in Table 3.

The common-dimension effect was strong in the equal and unequal-units conditions, contrary to the expectation that it would be easier to use the unique information in the equal-units condition, since there was less data transformation for the judge to consider. Another surprising result was the fact that the warning given to the informed groups failed to produce a significant reduction in the common-dimension effect. The warning did have one interesting consequence, however. The average multiple correlation between the predictions of Equations 2, 3, and 4 and the subjects' quantitative judgments was lower when the subjects had been informed about the common-dimension effect ($\bar{R} = .83$) than when they had not been so informed ($\bar{R} = .86$). Though small, the effect was statistically significant, $F(1,116) = 5.6$, $p < .05$. It appears that, instead of reducing the common-dimension bias, the warning made subjects less consistent in their use of information.[3]

Introspective comments. After they had finished, subjects were asked to write a brief description of how they had made their decisions. Although the general results of the experiment reflect certain difficulties subjects have in integrating information, the most interesting fact about their written protocols is that subjects made virtually no mention of any such difficulties. They rarely mentioned giving the common dimension any special consideration, though a few said they looked at it first. Instead, they typically indicated that they took each factor into proper account when making their decisions.

The protocols, written after the experiment, are obviously incomplete. It would seem worthwhile to obtain more systematic introspective data during the actual judgment period in order to examine the process strategies underlying the common-dimension effect.

Experiment II

A second study was conducted to test whether the common-dimension effect found in Experiment I could be due to some aspect of the stimuli other than the commonality built in by the paired-comparison design. A new group of subjects participated for two sessions. In Session 1 these subjects rated singly the prospective GPA for each of the 180 students depicted in the 90 pairs used in Experiment I. In Session 2, one day later, they judged the regular group of 90 pairs, exactly as was done in Experiment I.

The ratings of individual students from Session 1 were analyzed by making them into pseudo pairs. For example, if Student 27 and Student 161 from the set of 180 students had appeared as one of the 90 pairs in Experiment I, the difference between the two individual ratings of these students was used as a pseudo paired comparison in the analysis of data from Session 1. Thus, if the common-dimension effect in Experiment I was due to some aspect of the stimuli other than commonality, the effect should occur in the analysis of these pseudo pairs as well. If the effect is inherent only in real comparative judgments, it should not occur in these pseudo pairs.

Method

The task. During Session 1 subjects predicted the freshman GPA for each of the 180 students by putting a slash mark on the line scale shown at the bottom of Figure 1. The slash lines were later calibrated into a numerical score for purposes of analysis. During Session 2, the task was identical to that of Experiment I (see top of Figure 1).

Subjects. The subjects were 60 paid volunteers from the University of Oregon. Thirty subjects, 15 men and 15 women, were assigned to each of two experimental groups. Group 1 judged the equal-unit stimuli of Experiment I in both sessions. Group 2 judged the stimuli with unequal units. None of the subjects were informed about the common-dimension bias when they judged actual pairs during Session 2.

Data analysis. Once the ratings from Session 1 were made into pseudo comparisons, they were analyzed exactly as were the data in Experiment I. The data from Session 2 were also analyzed in the same way.

Results

The results from Experiment II are shown in the bottom half of Table 3. There was no enhanced importance of the common dimension evident in the analysis of the pseudo pairs. However, when these same subjects judged the real pairs during Session 2, dimensions were given more weight when they were common than they received when they were unique, as in Experiment I. The common-dimension effect based on the quantitative comparisons within the real pairs was statistically significant, $F(1,58) = 12.2$, $p < .001$, and the difference in the relative weighting of common and unique dimensions between the pseudo pairs and the real pairs was also significant, $F(1,116) = 9.3$, $p < .01$.

The remainder of the analyses of the real pairs in Experiment II also replicated the findings from Experiment I. Again, the common-dimension effect was equally strong in the analysis of the binary choices and it was not significantly affected by the stimulus units or by the type of common dimension.

The fact that the common-dimension effect occurred only with real pairs indicates that the judged difference between students cannot be accounted for by the difference between the judgments of the individual students. The results of Experiment II thus insure that the common-dimension effect is a contextual bias, inherent in the choice process itself, and not an artifact of some uncontrolled element in the individual stimuli.

Additional analyses. As described earlier, the primary method of data analysis fitted Equations 2, 3, and 4 to the responses from each subject.

The twelve weights from these three equations were then combined in a way that compared the weight that a dimension received when it was common to both stimuli with the weight that it received when it was unique. Although this is a meaningful way to summarize the data, several other analyses are also of interest. For example, a natural question is whether the common dimension generally received more weight than the two unique dimensions that accompanied it in the pair. To answer this question we must examine directly the twelve weights from Equations 2, 3, and 4. These weights are shown in Table 4, based on the real paired comparisons in Experiment II. It is evident that the weightings for the common dimension were not always higher than the weights given the unique dimensions in a pair. For example, Quant always received the highest weight in the unequal-units condition, regardless of whether it was a common or unique dimension, although it was less dominant when it was unique.

Insert Table 4 about here

Another perspective on the common-dimension effect is provided by Table 5, which illustrates, for three pairs, the proportion of times a student with a substantial advantage on the common dimension was chosen over a student who was superior on the unique dimension. Table 5 also provides a comparison, in Experiment II, between the choice frequencies and the orderings based on the ratings of the individual stimuli in the pair. For all three of these pairs, the student who was superior on the common dimension fared better when the response was a choice than when it was an individual rating. However, the student with an advantage on the common dimension was not always chosen and, as Pair II illustrates, might be chosen less than half of the time if the other student had a particularly good score on the unique dimension.

Insert Table 5 about here

In sum, the additional analyses depicted in Tables 4 and 5 indicate that the common dimensions, though more important than the unique dimensions, did not dominate the choices completely.

Experiment III

Experiments I and II have demonstrated that a dimension receives higher weight when it is common than when it is unique. It is important to note, however, that the subjects in those experiments were run in large groups and were given neither feedback nor explicit rewards for making careful decisions. In addition, the subjects received relatively little information about the distributional properties of the test scores. There is some question, therefore, whether the results will generalize to conditions in which greater degrees of information and motivation prevail.[4] Experiment III addresses this concern. In this experiment, the subjects were tested in relatively small groups. In addition, they were given "correct-answer" feedback after each judgment, their salary for the experiment was dependent upon the accuracy of their responses, and they were provided with detailed information about the distributions of _NAch_, _Eng_, _Quant_, and the distribution of correct answers.

Method

The task. The subjects evaluated comparative GPA as in Experiment I and Day 2 of Experiment II. Only the unequal-units stimuli were used and none of the subjects were informed in advance about the common-dimension bias. Instead of putting a slash mark on a line to indicate relative GPA, as in the earlier experiments, subjects wrote down the letter name of the superior student and made a direct numerical estimate of the difference (e.g., A12, B18, etc.) on a scale calibrated from 1 to 60.

Information and feedback. In addition to the mean scores for NAch, Eng, and Quant, subjects were given graphs showing the frequency distributions of these scores across the 90 pairs of stimuli.

After the subject made his response, he was provided with feedback about the "correct answer" (e.g., A14, B16, etc.). The correct answers were generated using a linear equation in which each score, whether common or unique, had a beta weight of .5. Thus, any degree of learning would reduce the common-dimension effect and a subject who was perfectly accurate would show no effect at all. A frequency distribution of the correct answers was provided and the subjects were given 15 practice trials with feedback prior to beginning the test trials.

Payment for accuracy. Subjects were told that they would receive a base payment of $1.00 for participating in the experiment and that the remainder of their salary would depend on the sum of the differences between their answer and the correct answer for eight selected pairs of stimuli. They were given a chart showing the relationship between accuracy and payment. Rewards for accuracy ranged between $0 and $4.

Subjects. The subjects were 30 volunteers, approximately equal numbers of men and women, from the University of Oregon. They were tested in groups of about seven to nine members.

Results

The results from Experiment III are shown in the bottom row of Table 6. There was again a statistically significant common-dimension effect, $F(1,29) = 13.6$, $p < .001$. Comparison with the comparable unequal-units and uninformed-about-bias conditions from Experiments I and II indicates that the effect in Experiment III was not in the least diminished by the provision of feedback, reward for accuracy, and distributional information. An analysis of variance showed no significant differences among the three experiments.

Insert Table 6 about here

Experiment IV

Experiments I, II, and III have demonstrated that a dimension receives a higher weight when it is common than when it is unique. Experiments IV and V were designed to probe more deeply into the nature and causes of this effect.

Amos Tversky[4] has proposed a Random Weight Model to account for the common-dimension effect. Under this model, it is hypothesized that the subject applies one weight to the difference between the stimuli on the common dimension and combines the resultant with the products of weights applied separately to each of the unique dimensions. Thus Equation (2) would be replaced by the following:

$$\text{Set N: } d_1(AB) = n_1(N_A - N_B) + e_3E_A - e_4E_B \qquad (5)$$

where n_i, e_i, and q_i are random variables. Similar changes would be made for Equations (3) and (4). An important feature of this model is the assumption that the weighting parameters are not fixed but, instead, are distributed about their central values. The application of each weight by the subject is assumed to be analogous to sampling a value at random from this distribution. The resultant variability in the application of the weights is intended to represent inherent noise or imperfection.[5] In contrast to independent evaluations of Stimuli A and B, which would require weighting the common dimension twice, the application of just one weight to the difference on the common dimension is a natural simplification of the task, similar to the behavior shown by subjects in an experiment by Tversky (1969).

For purposes of illustration, assume that $\underset{\sim}{n}_1 = \underset{\sim}{q}_3 = \underset{\sim}{e}_4$ in Equation (5) and that the variances of the distributions of these weights are all equal. If the data resulting from these and the random sampling assumptions were analyzed by the approach used in Experiments I, II, and III, the two weights estimated for the common dimension would, on the average, exceed the weights estimated for the two unique dimensions. Since the weight for the common dimension is sampled only once, while the weights for the unique dimensions are sampled twice, more error becomes associated with the unique dimensions. This increase in error leads to smaller weights for these dimensions. In other words, the Random Weight Model introduces a subtle bias in favor of the common dimension. Even though the subject was, on the average, applying equal weights to all dimensions, the common dimensions would correlate more highly with his responses across a series of judgments.

One implication of the Random Weight Model can be tested using the weights estimated in Experiment II. If the model is valid, the two weights calculated for the common dimension in the real-pair condition (one weight for Stimulus A, the other for Stimulus B) should be more similar to one another than the two weights calculated for that same dimension in the pseudo-pairs condition. In the latter condition, the subject judged Stimulus A and Stimulus B separately and would have had to weight the common dimension each time, thus contributing the error variability twice, while in the real-pair condition the model assumes that only one weight is applied.

Table 7 presents the mean differences between weights for the common dimension as a function of the type of stimulus pair (real vs. pseudo) being judged. The statistical significance of these data was evaluated by means of an analysis of variance done separately for the equal and unequal-units conditions. In the equal-units condition, the difference between weights was larger for pseudo pairs than for real pairs, $F(1,29) = 5.4$, $p < .05$, thus providing some support for the Random Weight Model. However, this effect was not found in the unequal-units condition. In addition, the difference between pseudo and real pairs on the overall mean for both conditions was not significant. In summary, while there was not unanimous agreement between the prediction of the Random Weight Model and the data, there was enough agreement to suggest that the model may have some validity.[6]

Insert Table 7 about here

Because of the possibility that the common-dimension effect found in Experiments I, II, and III could have been due to a process analogous to that suggested by the Random Weight Model, Experiment IV was designed to test for the effect with a design that was not susceptible to random error in weighting. The finding of a common-dimension effect with such a design would not exclude the Random Weight Model, but it would indicate that something other than a random-weighting procedure was increasing the salience of the common dimension.

Method

The task. The basic task again involved the evaluation of freshman GPA on the basis of NAch, Eng, and Quant. The judgments about individual students made during Session 1 of Experiment II served as a basis for selecting pairs of students with approximately equal ratings. Subjects in Experiment IV were asked to make comparative judgments about these pairs of students, following the same instructions used in Experiments I and II. The common-dimension hypothesis predicts that subjects would exhibit a definite preference for the student who had a superior score on the common dimension. Use of this design avoids dependence upon linear equations whose weights are susceptible to the influence of random error.

Stimuli. Both equal and unequal-units conditions were included in this study. Within each condition the mean evaluations of each of the 180 individual stimuli from Experiment II were used as a basis for equating pairs of students to be compared in the present study. Twenty-nine pairs were constructed for the equal-units condition and 28 pairs for the unequal-units condition. The three scores, NAch, Eng, and Quant, served as the common dimension about equally often. Five practice pairs and enough filler pairs to bring the total number of pairs to 50 were included in the set seen by subjects in each condition.

The equating of stimuli within a pair was quite good. For the 28 unequal-units pairs, the average absolute difference in mean rating for each stimulus was 0.15 units on a scale in which the range of individual means was from 21.0 to 44.4 The stimulus that was superior on the common dimension was given the higher individual rating 49.3% of the time and had a mean rating of 0.04 units less than the stimulus with which it was paired. Comparing each pair of stimuli across the 30 subjects who rated the stimuli in Experiment II, the dominant stimulus had a higher rating for an average of 15.3 subjects, the other was higher for 12.9 subjects, and 1.7 were tied. In short, the stimuli within each pair of the unequal-units condition were exceptionally well matched on the basis of the single-stimulus ratings of Experiment II. The stimuli in the equal-units condition were matched just as closely.

Subjects. The subjects were 64 paid volunteers from the University of Oregon. Thirty-four of these participated in the equal-units condition and the remainder participated in the unequal-units condition. Approximately equal numbers of men and women were assigned to each condition.

Results

Unequal-units condition. The alternative that was superior on the common dimension was selected in 58.3 percent of the 835 choices made by 30 subjects for the 28 equated pairs (the subjects failed to choose on five occasions). The alternative higher on the common dimension was selected more often than the other alternative on 19 pairs, less often on 7 pairs, and equally often on 2 pairs ($p < .025$ according to the binomial test). Thus the common dimension had more influence than unique information in a situation where estimation of influence was not confounded with error variability in the weighting process.

Equal-units condition. Across the 29 matched pairs, the stimulus that was superior on the common dimension was selected 63% of the time (it was selected more often than the other stimulus for 22 of those pairs). However, there was an alternative to the common-dimension hypothesis that was able to predict choices almost as well. It was suggested by the post-experiment comments of several subjects who said that they weighted all four test scores equally in the comparison. They simply summed the two scores for each student and chose the student with the higher sum. Examination of the data showed that the student with the higher sum was selected 61.7% of the time. For pairs in which the student who was superior on the common dimension had a lower sum, that student was selected only 51% of the time. When the student who was superior on the common dimension also had a higher sum, that student was selected 71% of the time--a value that is about equal to a combination of the main effects due to sum and common dimension. It appears that both the sum and the common dimension are potent determiners of choices in the equal-units condition.

Conclusion. Although random error in weighting cannot be ruled out as a contributor to the effects found in the first three experiments, it appears unlikely that it is the sole cause of those effects. In Experiment IV, a design that was not susceptible to influence from random error in weighting also indicated increased salience for common dimensions.

Experiment V

If the common-dimension effect is not simply the result of random error in weighting, what are the cognitive mechanisms by which it occurs? Experiment V was designed to test the hypothesis that subjects weight common dimensions more because they attend first to common information, form an initial impression based on this information, and then fail to adjust this impression sufficiently when they attend to the unique information. We shall refer to this as the "adjustment hypothesis."

A priori support for the adjustment hypothesis comes from several sources. Lichtenstein and Slovic (1971) found evidence that subjects evaluating gambles formed an initial impression on the basis of a single attribute and then failed to adjust this impression sufficiently when considering the other attributes of the gamble. The first impression thus carried the most weight. In addition, the adjustment hypothesis leads to a primacy effect of the sort found in many studies where subjects had to integrate information sequentially prior to making a judgment (see, for example, Slovic & Lichtenstein, 1971, pp. 692-693).

Method

Stimuli. The stimuli were the first 25 pairs of students from the set of 90 pairs judged in the unequal-units condition of Experiments I and II. As before, the same instructions and procedures applied with one important exception. The test scores were presented sequentially rather than simultaneously. One group of subjects always saw the two common scores for Students A and B first. They were instructed to form an initial impression on the basis of these scores and then turn the page to receive the two unique scores, after which they made their response. A second group of subjects saw the two unique scores first, then the scores on the common dimension. If the adjustment hypothesis is correct, then the average preference for the student who was superior on the common dimension should be greater when the common scores are given first.

<u>Subjects</u>. The subjects were 21 men, paid volunteers from the University of Oregon. Eleven of these saw common information first, the remainder received unique information first.

<u>Results</u>

The student who was superior on the common dimension was selected 69% of the time when the common scores were presented first and 74% of the time when the unique scores came first. This difference was not statistically significant. Thus the information attended to first was not weighted more heavily in the comparison, contrary to the adjustment hypothesis.

Discussion

The present experiments indicate that the property of providing information for both alternatives of a pair enhances the degree to which a stimulus dimension is weighted in the comparison between those alternatives. This result is important in several ways. First, it is congruent with a growing body of experimental literature that documents man's difficulties in weighting and combining, or in other ways transforming, information. Second, it has implications for real-world decision making. These and related topics will be discussed below.

Explanations for the effect. The general presumption underlying these experiments was that use of the common dimension would be facilitated by the fact that this dimension provides a direct and unambiguous comparison between the alternatives on the attribute being judged. To use unique information, one must deal with questions of relative weights or "tradeoffs" between different attributes. It seemed likely that the cognitive difficulty of estimating and implementing such tradeoffs would decrease the use of unique information. Although the common-dimension effect has now been demonstrated, the question of whether "cognitive-strain" provides a satisfactory explanatory framework remains open. Only one direct manipulation of strain was attempted, the use of equal and unequal-units conditions in Experiments I, II, and IV. Contrary to expectation, the common-dimension effect was not greater in the unequal-units conditions, the conditions that should have been most difficult to judge.

Although several other hypotheses were advanced and tested in an attempt to explicate the common-dimension effect, the basic mechanisms remain unclear. Some support was found for a Random Weight Model which postulated that the lesser importance of unique information was due to greater random error in the use of that information. However, strong direct support for this model is lacking. In addition, the results of Experiment IV indicate that there must be more to the effect than random error in the application of the weights, although the attempt in Experiment V to demonstrate that the effect was due to a starting-point and adjustment process was not successful.

One possible explanation for the effect, not tested here, is that subjects might actively discount the unique information because they lack confidence in their ability to use it appropriately. In other words, the ease of comparison on the common dimension could induce greater confidence in one's ability to use that information, and this confidence, in turn, could mediate the weight given the information.

John Castellan[6] has suggested yet another explanation. He hypothesized that subjects may substitute the average scores for cue dimensions that are missing. Substitution of average scores for missing values implies that the intradimensional difference between stimuli will be larger on the common dimensions than on the unique dimensions. Tversky (1969) has found that subjects tend to discount dimensions on which stimuli have small differences. If subjects did that here, a bias in favor of the common dimension would occur. While this hypothesis is untested, its plausibility is enhanced by the introspective comments of a number of subjects who volunteered that they did tend to assume average values for missing scores and did discount small differences between stimuli on a dimension.

Information transformation and information use. Slovic and Lichtenstein (1968a) observed a "compatibility" effect in judgments of single stimuli that is not unlike the common-dimension effect in the present study. They found that when subjects rated the attractiveness of a gamble, probability of winning was the most important factor in their policy equations. In a second condition, subjects were required to indicate the attractiveness of a gamble by an alternative method--namely equating the gamble with an amount of money such that they would be indifferent between playing the gamble and receiving the stated amount. Here it was found that attractiveness was determined more by a gamble's potential outcomes than by its probabilities. The outcomes, being expressed in units of dollars, were readily commensurable with the units of the responses--also dollars. This commensurability led subjects to use one of the outcomes as a starting point for the response and these starting points became the primary determiners of the responses via the mechanism of insufficient adjustment discussed in conjunction with Experiment V. On the other hand, the probability cues had to be transformed by the subject into values commensurable with dollars before they could be integrated with these other cues. It seems plausible that the cognitive effort involved in making this sort of transformation greatly detracted from the influence of the probability cues in the second task.

In the area of comparative judgment, empirical work further supports the general conclusion that subjects find it rather difficult to weight and "trade off" values in a compensatory manner. For example, Tversky (1969) observed systematic intransitivities in subjects' choices among pairs of gambles described by probability of winning and amount to win. He explained these intransitivities in terms of a "lexicographic semiorder" decision structure whereby subjects ignored small differences in probability and chose among adjacent gambles on the basis of the payoff value. However, when gambles were further apart in the chain, subjects chose according to probability or expected value. In sum, Tversky's subjects did not exhibit a uniform trade-off between the utility and probability dimensions. Shepard (1964) also found evidence for the difficulty of combining information in several perceptual judgment tasks.

Support seems to be building for a general hypothesis to the effect that information that has to be held in memory, inferred, or transformed in any but the simplest ways, will be discounted. The present study and the work just discussed are congruent with this hypothesis. Additional evidence for this notion comes from the tendency for subjects to prefer a direct test of any hypothesis they are trying to validate (Bruner, Goodnow, & Austin, 1954; Wason, 1968). Indirect tests require a transformation of information. Bruner et al. found that subjects preferred to make direct tests with positive instances and concluded that information that does not require "in the head" transformations is preferred in the interest of cognitive economy. Resistance to making simple transformations of information prior to using that information has also been found by Slovic and Lichtenstein (1968b) and Payne and Braunstein (1971) in the context of gambling decisions.

Implications. Experimental work documents man's difficulties in weighing and combining information. Do these difficulties diminish once the subject leaves the artificial confines of the laboratory and resumes the task of using familiar sources of information to make decisions that are personally important to him? While there is a lack of evidence bearing on this question, there are some hints, at least, that systematic biases pervade the judgments of experts in many areas, including medicine (Bakwin, 1945), industry (Kidd, 1971), and science (Tversky & Kahneman, 1971).

Future experimentation should examine whether the "common-dimension effect" and similar types of bias will pervade important decision situations. If they do, further work will be needed to determine how to minimize these biases. Education may help, although a casual warning to avoid the bias did not work in the present study. Substitution of formal systems for combining the decision maker's values (Dawes, 1971; Edwards, Lindman, & Phillips, 1965; Goldberg, 1970; Raiffa, 1968) may provide the necessary means for analyzing and eliminating biases.

References

Bakwin, H. Pseudodoxia pediatrica. New England Journal of Medicine, 1945, 232, 691-697.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Dawes, R. M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-188.

Edwards, W., Lindman, H., & Phillips, L. D. Emerging technologies for making decisions. New directions in psychology II. New York: Holt, Rinehart, & Winston, 1965. Pp. 261-325.

Goldberg, L. R. Man vs. model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.

Kidd, J. B. The utilization of subjective probabilities in production planning. Acta Psychologica, 1970, 34, 338-347.

Lichtenstein, S., & Slovic, P. Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology, 1971, 89, 46-55.

Payne, J. W., & Braunstein, M. L. Preferences among gambles with equal underlying distributions. Journal of Experimental Psychology, 1971, 87, 13-18.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

Shepard, R. N. Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1964, 1, 54-87.

Slovic, P., & Lichtenstein, S. C. The relative importance of probabilities and payoffs in risk taking. Journal of Experimental Psychology Monograph Supplement, 1968, 78, No. 3, Part 2. (a)

Slovic, P., & Lichtenstein, S. C. The importance of variance preferences in gambling decisions. Journal of Experimental Psychology, 1968, 78, 646-654. (b)

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Tversky, A. Intransitivity of preferences. Psychological Review, 1969, 76, 31-48.

Tversky, A. Choice by elimination. Journal of Mathematical Psychology, 1972, 9, 341-367.

Tversky, A., & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.

Wason, P. C. On the failure to eliminate hypotheses--a second look. In P. C. Wason & P. M. Johnson-Laird (Eds.), Thinking and reasoning. Baltimore: Penguin Books, 1968. Pp. 165-174.

Footnotes

[1]Sponsorship for this research comes from the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-68-C-0431, Contract Authority Identification No. NR-153-311; from Grants MH12972, MH15414, and MH21216 from the U.S. Public Health Service; and Grant GS-32505 from the National Science Foundation. Computing assistance was obtained from Health Sciences Computing Facility, U.C.L.A., sponsored by NIH Grant RR-3.

We are indebted to John Castellan, Robyn Dawes, Lewis Goldberg, Daniel Kahneman, Sarah Lichtenstein, and Amos Tversky for their many helpful suggestions regarding the design and reporting of this study.

[2]Requests for reprints should be sent to Paul Slovic, Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

[3]An alternative hypothesis is that the warning led subjects to apply systematic strategies of the sort that would not be detected by the model of analysis used in the present study.

[4]Personal communication.

[5]For a formal discussion of a more general class of Random Weight Models see Tversky (in press).

[6]Personal communication.

Table 1

Means and Standard Deviations

for the Stimulus Dimensions

	Dimension	Mean	Standard Deviation
Unequal-Units Condition	NAch	50	20
	Eng	100	20
	Quant	500	150
Equal-Units Condition	NAch	500	150
	Eng	500	150
	Quant	500	150

Table 2

Examples of Stimulus Pairs in the

Equal- and Unequal-Units Conditions

Common Dimension	Unequal-Units Condition			Equal-Units Condition		
	Student →	A	B	Student →	A	B
NAch	NAch	67	59	NAch	618	561
	Eng	___	86	Eng	___	572
	Quant	452	___	Quant	382	___
		A	B		A	B
Eng	NAch	___	33	NAch	___	458
	Eng	119	90	Eng	457	800
	Quant	414	___	Quant	348	___
		A	B		A	B
Quant	NAch	___	27	NAch	___	469
	Eng	74	___	Eng	698	___
	Quant	701	466	Quant	264	388

Table 3

Mean Weights for Common and Unique Dimensions

Experimental Condition	Quantitative Comparison		Binary Choice	
	Common Dimensions	Unique Dimensions	Common Dimensions	Unique Dimensions
EXPERIMENT I:				
Overall Mean	.83	.73	.73	.64
Uninformed	.85	.74	.74	.64
Informed	.81	.72	.71	.64
Unequal Units	.82	.71	.72	.62
Equal Units	.84	.76	.74	.67
NAch Common	.81	.71	.74	.64
Eng Common	.79	.68	.68	.61
Quant Common	.89	.81	.76	.68
EXPERIMENT II: PSEUDO PAIRS				
Overall Mean	.80	.80		
Unequal Units	.80	.79		
Equal Units	.81	.80		
EXPERIMENT II: REAL PAIRS				
Overall Mean	.85	.72	.73	.64
Unequal Units	.80	.71	.71	.60
Equal Units	.89	.74	.76	.67
NAch Common	.78	.67	.73	.64
Eng Common	.85	.71	.69	.60
Quant Common	.91	.81	.78	.67

Table 4

Mean of the Absolute Weights for Decisions about Real Pairs: Experiment II

	Set N				Set E				Set Q			
Experimental Condition	n_1	n_2	q_3	e_4	e_1	e_2	q_4	n_3	q_1	q_2	e_3	n_4
Quantitative Comparison												
Unequal Units	.40	.34	.48	.30	.37	.37	.42	.34	.43	.48	.26	.32
Equal Units	.42	.40	.39	.43	.43	.53	.33	.32	.46	.44	.42	.36
Binary Choice												
Unequal Units	.40	.30	.35	.29	.35	.32	.32	.27	.28	.48	.28	.29
Equal Units	.36	.31	.40	.34	.36	.44	.26	.33	.39	.42	.38	.32

Table 5

Ratings and Choices for Selected Pairs of Stimuli in Experiment II

	A		B	
Pair I	NAch	___	NAch	667
	Eng	491	Eng	___
	Quant	742	Quant	342

Pair I		Higher Rating A	Higher Rating B	
Choice	A	21	8	29
	B	1	0	1
		22	8	

	A		B	
Pair II	NAch	___	NAch	474
	Eng	470	Eng	566
	Quant	674	Quant	___

Pair II		Higher Rating A	Higher Rating B	
Choice	A	14	2	16
	B	10	0	10
		24	2	

	A		B	
Pair III	NAch	___	NAch	30
	Eng	90	Eng	131
	Quant	602	Quant	___

Pair III		Higher Rating A	Higher Rating B	
Choice	A	10	1	11
	B	11	6	17
		21	7	

Note--Pairs I and II were taken from the equal units condition; Pair III from the unequal units condition. The cell entries in the 2 × 2 tables indicate the number of subjects who chose the student designated by the row letter and gave a higher rating to the student designated by the column letter.

Table 6

Mean Weights for Common and Unique Dimensions

in Experiments I, II, and III

	Overall Mean		
	Common Dimensions	Unique Dimensions	% Ss Showing Effect
Experiment I	.85	.73	77%
Experiment II (Day 2 - Real Pairs)	.80	.71	63%
Experiment III (Feedback and Reward)	.89	.77	77%

Note:--Only the data from the uninformed groups, unequal-units conditions of Experiments I and II are presented here.

Table 7

Comparison of the Mean Difference Between Weights for the Common Dimension in Real and Pseudo Pairs of Experiment II

	Pseudo Pairs	Real Pairs
Equal-Units Condition	.149	.117
Unequal-Units Condition	.120	.118
Overall Mean	.135	.118

Note.--The dependent variable for this analysis is the absolute difference between the two weights for the common dimension.

Figure Caption

Figure 1. Response scales used in Experiments I and II. In Experiment I (top scale), subjects circled the letter corresponding to their judgment and estimated the difference between the students by placing a slash mark on the line scale. In Session 1 of Experiment II (bottom scale), subjects used the slash mark to estimate GPA for individual students.

ACADEMIC PRESS, INC.
PUBLISHERS
111 FIFTH AVENUE · NEW YORK, N. Y. 10003 · TELEPHONE OREGON 7-6713
CABLE ADDRESS: ACADEMIC NEWYORK • TELEX: DOM. 12-7081 • INTL. 235595

NOTICE TO AUTHORS

1. ORIGINAL FIGURES SHOULD BE RETURNED ONLY IF CORRECTIONS ARE NECESSARY. In such a case, please follow the instructions below.

   Indicate the correction on the engraver's proof, NOT ON THE ORIGINAL DRAWING OR GLOSSY PHOTOGRAPH.

   Most corrections will require a new engraving of the entire figure. Therefore, a draftsman should correct the original art by using the same size and style of lettering that was used to draw the figure initially.

   Return the corrected art and engraver's proof with the galleys.

2. HALFTONE REPRODUCTION

   The reproduction of halftones should be judged only from the proofs supplied on coated paper.

3. OVERSIZE FIGURES

   To make an engraving of a drawing larger than $8\frac{1}{2}$ x 11 inches, a reduced photoprint occasionally will be used. In such cases, the original, oversize drawing will not be returned unless requested, and will be discarded two months after publication of the article.

ACADEMIC PRESS, INC.
Journal Production Department

OBHP-386-1 (40)

Difference Between Students

Better Student

Pair 1. A (B)

No Difference

Very Large Difference

GPA

Student No. 1

0 F

1 D

2 C

3 B

4 A

OBHP-386-1

40

Table 3

Mean Weights for Common and Unique Dimensions

Experimental Condition	Quantitative Comparison		Binary Choice	
	Common Dimensions	Unique Dimensions	Common Dimensions	Unique Dimensions
EXPERIMENT I:				
Overall Mean	.83	.73	.73	.64
Uninformed	.85	.74	.74	.64
Informed	.81	.72	.71	.64
Unequal Units	.82	.71	.72	.62
Equal Units	.84	.76	.74	.67
NAch Common	.81	.71	.74	.64
Eng Common	.79	.68	.68	.61
Quant Common	.89	.81	.76	.68
EXPERIMENT II: PSEUDO PAIRS				
Overall Mean	.80	.80		
Unequal Units	.80	.79		
Equal Units	.81	.80		
EXPERIMENT II: REAL PAIRS				
Overall Mean	.85	.72	.73	.64
Unequal Units	.80	.71	.71	.60
Equal Units	.89	.74	.76	.67
NAch Common	.78	.67	.73	.64
Eng Common	.85	.71	.69	.60
Quant Common	.91	.81	.78	.67