Trade-Offs Depend on Attribute Range

Barbara A. Mellers and Alan D. J. Cooke

Most judgments depend on multiple attributes. Assessments of jobs often depend on salary level, potential for advancement, size and location of the firm, and vacation time. Impressions of potential marriage partners might depend on sense of humor, intelligence, physical attractiveness, and financial stability. Evaluations of consumer products, such as cars, might depend on price, gas mileage, comfort, and style. These complex judgments often involve trade-offs. How much compensation will an employee give up for two extra weeks of vacation? How much more will a consumer pay for a car to get 10 more miles per gallon? Because our lives so often involve trade-offs, it is important that descriptive theories of judgment incorporate mechanisms by which these trade-offs can occur.

Multiattribute judgments can be represented as a composition of functions (Birnbaum, 1974a; Massaro & Friedman, 1990), as in Figure 1. To illustrate with a simple example, imagine a consumer who is considering cars on the basis of price and gas mileage. In the evaluation stage, physical values, \( \phi_{\text{price}} \) and \( \phi_{\text{gas}} \) are converted to their corresponding subjective values, \( s_{\text{price}} \) and \( s_{\text{gas}} \) by means of psychophysical functions, denoted \( H \). In the integration stage, scales are combined by means of a composition rule, \( C \), to form an overall impression of the car, \( \Psi \). In the decision stage, the overall impression is translated to an overt response, \( R \), through a judgment function, \( J \).

Many experiments have shown that multiattribute judgments can be described by some form of averaging (Birnbaum & Stegner, 1979, 1981; Nisbett, Zukier, & Lemley, 1981; Surber, 1981; Troutman & Shanteau, 1977). In some cases, judgments are consistent with a relative-weight averaging model (Birnbaum, 1976; Birnbaum, Wong, & Wong, 1976), although systematic deviations from this model have been found (Birnbaum, 1974a; Birnbaum & Stegner, 1979, 1981; Fiske, 1980; Ronis & Lipinska, 1985). According to the relative-weight averaging model, the judged attractiveness of a car on the basis of gas mileage and price is expressed as

\[
R_{jk} = \left( w_p s_{\text{price}}^j + w_g s_{\text{gas}}^k + w_i \right) / \left( w_p + w_g + w_i \right),
\]

where \( w_p \), \( w_g \), and \( w_i \) are weights reflecting the subjective importance of price, gas mileage, and the initial impression, respectively; \( s_{\text{price}} \) and \( s_{\text{gas}} \) are the subjective values of price, gas mileage, and the initial impression. The initial impression represents the psychological value of a car prior to any specific information, somewhat analogous to a prior probability (Kaplan, 1971).

The relative-weight averaging model is one mechanism by which trade-offs can occur. Increasing values along one attribute can compensate for decreasing values on another. For example, a consumer might be indifferent between a car that costs $22,000 that gets only 20 miles per gallon (mpg) and a more expensive car priced at $25,000 that gets 30 mpg.

Factors That Influence Trade-Offs

Trade-offs between attributes are assumed to vary when the rank orders of judgments assigned to the same stimuli differ across conditions. Several factors have been shown to produce rank order shifts. A large literature on preference reversals demonstrates that changes in the response mode can alter the preference order of gambles (Birnbaum & Sutton, 1992; Goldstein & Einhorn, 1987; Grether & Plott, 1979; Mellers, Chang, Birnbaum, & Ordoñez, 1992; Mellers, Ordoñez, & Birnbaum, 1992; Slovic & Lichtenstein, 1983; Tversky, Sattath, & Slovic, 1988). A gamble with a large probability of winning a small amount is often
judged more attractive than one with a small probability of winning a large amount, but the latter gamble is assigned a higher selling price than the former. Framing effects have also been shown to influence the rank order of preferences (Hershey & Schoemaker, 1980; Kahneman & Tversky, 1984; McNeil, Pauker, Sox, & Tversky, 1982). One social policy is preferred over another when programs are described in terms of lives saved, but the other policy is preferred when the same two programs are described in terms of lives lost. Furthermore, changes in the judge's point of view (e.g., from buyers to sellers) yield systematic shifts in rank orders (Birnbaum, Coffey, Mellers, & Weiss, 1992; Birnbaum & Stegner, 1979). Sellers want more to sell a gamble with a 50% chance of winning either $96 or $0 than a gamble with a 50% chance of winning either $48 or $42, but buyers pay more to play the latter gamble than the former. Finally, contextual effects due to variations in stimulus spacing influence the rank order of judgments (Mellers, 1982; Mellers & Birnbaum, 1982). In the present article we examined another aspect of the context—attribute range. Does the range of an attribute influence the trade-offs people make in multiattribute judgments?

Range Effects

Numerous studies have demonstrated range effects with single attributes using category ratings (Parducci, 1965; Parducci & Perrett, 1971), magnitude estimations (Poulton, 1968; Teghtsoonian, 1971), and absolute judgments (Gravetter, 1971; Gravetter & Lockhead, 1973; Holland, 1968). Results from these studies show that a given attribute difference has a greater effect in a narrow range than in a wide range. For example, suppose consumers judge the attractiveness of cars on the basis of price only, and the range of price is varied across contexts. One context has a narrow price range, and the other has a wider price range. Price changes for the cars common to both ranges have a greater effect on attractiveness ratings in the narrow range than in the wide range.

This pattern of results is predicted by Parducci's (1968, 1974) range–frequency theory. Range–frequency theory asserts that single-attribute judgments are a compromise between two principles. The range principle is a tendency to spread the range of stimuli across the entire range of the response scale, and the frequency principle is a tendency to use each response with equal frequency. Holding stimulus frequency constant, range–frequency theory implies that changes in the attribute range should produce linear response changes, and responses to common stimuli in the narrow range should span a wider interval than responses in the wide range.

Loci of Range Effects

In the present article we investigate the loci of range effects in multiattribute judgments. We entertain three loci, represented by the functions in Figure 1.

Response effects. Range effects could be semantic and influence the use of the response scale. In this case, range effects would be attributed to the output transformation (\(J\) in Figure 1). If variations in attribute range influence the response but not the perception or the weighting of the stimulus, range effects would be attributed to the judgment function (Hutchinson, 1983; Lynch, Chakravarti, & Mitra, 1991). Because response effects are typically assumed to be monotonic, ordinal changes in judgments assigned to the same stimuli in different range contexts are inconsistent with this hypothesis. Mellers and Birnbaum (1982) found that the rank order of stimuli varied as a function of range and spacing, contrary to this response scale hypothesis.

Perceptual effects. An alternative possibility is that range effects are perceptual; they alter the mapping of physical value into subjective value (\(H\) in Figure 1). If attribute range influences the perception of the stimulus, the rank orders of the judgments would tend to vary. A scale–change hypothesis implies that the subjective values of attribute levels common to different ranges vary. That is, a given level of an attribute is psychologically different in one range compared with another.

Weighting effects. A third possibility is that range effects alter the subjective importance of an attribute. Such weighting effects would be attributed to changes in \(w\) in Equation 1 (when \(C\) in Figure 1 is an averaging process). Weight effects imply that the relative importance of an attribute varies systematically with range.

Figure 2 illustrates how shifts in the rank orders of the judgments could occur from either changes in weights or changes in scales. Rank orders of attractiveness ratings predicted from the relative-weight averaging model are presented in a baseline condition (Panel A) and two comparison conditions (Panels B and C). Arrows show the direction of preference for pairs of cars; one car has better gas mileage, and the other has a lower price.

In Panel A, arrows point down, indicating that cars with better gas mileage are more attractive than cars with lower prices. Changes from Panel A to Panel B are due to an increase in the weight of price; all other parameters are the same as those in Panel A. Those same rank order changes are reproduced in Panel C, where the range of scale values for price is increased, whereas all other parameters are the
range influences attribute weight. Marginal means of price (averaged over gas mileage) are plotted with a separate curve for each price range. According to the relative-weight averaging model (Equation 1), the slopes of these curves can be expressed as
\[ \Delta R = \Delta s_p w_p / (w_p + w_g + w_i), \]  
where \( \Delta R \) represents the change in marginal means, \( \Delta s_p \) is the change in price scale values, and \( w_p \), \( w_g \), and \( w_i \) are the weights (Equation 1). If the weight of price, \( w_p \), varies with attribute range, and the weight for the narrow context is larger than the weight for the wide context, the narrow range curve will be steeper than the wide range curve, as shown in Panel A. This pattern has been found with research on single-stimulus judgments, as discussed earlier.

If the range of price influences the subjective values of price, \( \Delta s_p \), and a given physical difference in price has a wider subjective interval in the narrow range context than in the wide range context, predicted curves will appear as in Panel B of Figure 4. The narrow range curve is steeper than the wide range curve.

The important point to note about Figure 4 is that Panels A and B are identical. Recall that price is varied across range contexts. Weight and scale hypotheses make equivalent predictions about the interaction between price and price range. To distinguish between weight and scale hypotheses, it is necessary to examine the interaction between gas mileage (an attribute with a constant range in both contexts) and price range, as shown in Figure 5.

Panel A shows predicted results if attribute range influences weight. Marginal means of gas mileage are plotted with a separate curve for each price range. According to the relative-weight averaging model, the slopes of these curves can be expressed as
\[ \Delta R = \Delta s_p w_p / (w_p + w_g + w_i), \]  
where \( \Delta s_p \) is the change in the subjective values for gas mileage. Notice that \( \Delta R \) is inversely related to changes in \( w_p \), because \( w_p \) is in the denominator. If price range influences the weight of price, the steeper curve in Figure 4 (narrow range) will be the flatter curve in Figure 5.

Panel B of Figure 5 shows predicted results if attribute range influences scales, and not weights. \( \Delta R \) in Equation 3 does not depend on \( \Delta s_p \). Slopes of the curves for the narrow and wide price ranges should be the same; there should be no interaction between gas mileage and price range. Slopes are parallel because \( \Delta R \) is unaffected by \( \Delta s_p \).

To summarize, the Price X Price Range interaction (Figure 4) is important to assess because it demonstrates the range effects and shows the direction of those effects. But this interaction does not discriminate between weight and scale shifts in rank order can result from differences in weights or differences in scales.

**Distinguishing Perceptual Effects From Weighting Effects**

It is possible to determine whether attribute range influences weights or scales in the context of averaging models. Birnbaum and Stegner (1979) developed a procedure for the relative-weight averaging model, and the process can be extended to other forms of averaging. To illustrate, suppose that consumers judge the attractiveness of cars on the basis of price and gas mileage, and the range of price is varied across contexts. One context has a narrow price range, the other has a wide price range, and the range of gas mileage is held constant in both contexts.

Panel A of Figure 4 shows predicted results if attribute range influences attribute weight. Marginal means of price (averaged over gas mileage) are plotted with a separate curve for each price range. According to the relative-weight averaging model (Equation 1), the slopes of these curves can be expressed as
\[ \Delta R = \Delta s_p w_p / (w_p + w_g + w_i), \]  
where \( \Delta R \) represents the change in marginal means, \( \Delta s_p \) is the change in price scale values, and \( w_p \), \( w_g \), and \( w_i \) are the weights (Equation 1). If the weight of price, \( w_p \), varies with attribute range, and the weight for the narrow context is larger than the weight for the wide context, the narrow range curve will be steeper than the wide range curve, as shown in Panel A. This pattern has been found with research on single-stimulus judgments, as discussed earlier.

If the range of price influences the subjective values of price, \( \Delta s_p \), and a given physical difference in price has a wider subjective interval in the narrow range context than in the wide range context, predicted curves will appear as in Panel B of Figure 4. The narrow range curve is steeper than the wide range curve.

The important point to note about Figure 4 is that Panels A and B are identical. Recall that price is varied across range contexts. Weight and scale hypotheses make equivalent predictions about the interaction between price and price range. To distinguish between weight and scale hypotheses, it is necessary to examine the interaction between gas mileage (an attribute with a constant range in both contexts) and price range, as shown in Figure 5.

Panel A shows predicted results if attribute range influences weight. Marginal means of gas mileage are plotted with a separate curve for each price range. According to the relative-weight averaging model, the slopes of these curves can be expressed as
\[ \Delta R = \Delta s_p w_p / (w_p + w_g + w_i), \]  
where \( \Delta s_p \) is the change in the subjective values for gas mileage. Notice that \( \Delta R \) is inversely related to changes in \( w_p \), because \( w_p \) is in the denominator. If price range influences the weight of price, the steeper curve in Figure 4 (narrow range) will be the flatter curve in Figure 5.

Panel B of Figure 5 shows predicted results if attribute range influences scales, and not weights. \( \Delta R \) in Equation 3 does not depend on \( \Delta s_p \). Slopes of the curves for the narrow and wide price ranges should be the same; there should be no interaction between gas mileage and price range. Slopes are parallel because \( \Delta R \) is unaffected by \( \Delta s_p \).

To summarize, the Price X Price Range interaction (Figure 4) is important to assess because it demonstrates the range effects and shows the direction of those effects. But this interaction does not discriminate between weight and scale shifts in rank order can result from differences in weights or differences in scales.
scale hypotheses. These hypotheses can be distinguished with another test—the interaction between gas mileage (a range-constant attribute) and price range (Figure 5). If price range influences the weight of price, there should not only be an interaction, but a very specific interaction: Slopes of the curves in Figure 5 should be inversely related to those in Figure 4. The steeper curve in Figure 4 should be the flatter curve in Figure 5. However, if price range influences the scale values, there should be no interaction. Slopes of the curves reflecting the effect of gas mileage should be identical.

Judged Weights

In the present article, weights were estimated in two ways. First, weights were estimated from the overall responses. This procedure is discussed in more detail later. We refer to these values as estimated weights. Second, subjects were asked to allocate 100 points among three attributes to reflect the relative importance of each. We refer to these responses as judged weights, as opposed to estimated weights. We were hopeful that subjects could articulate their decision strategies, and judged weights would be a good reflection of estimated weights derived from the fit of a model to the overall responses. If so, a weight-change hypothesis about attribute range might imply that judged weights should also vary. However, a scale-change hypothesis might be consistent with the null result—namely, judged weights would not vary with attribute range. We investigate judged weights in an attempt to provide behavioral implications of the weight-change hypothesis versus the scale-change hypothesis.

Do Some Attributes Show Greater Range Effects Than Others?

Beattie and Baron (1991) and Hutchinson (1983) suggested that range effects might be more likely to occur with some attributes than others. We examine this conjecture in three different experiments. Goldstein (1990) distinguished between global and local weights. Global weights are general indicators of a person's values and are largely independent of the particular stimuli involved. Local weights depend on both personal characteristics of the decision maker and the stimulus set. Local weights are only assumed to represent relative importance with respect to a particular stimulus set.

In this article, we distinguish between global and local ranges. A global range refers to the widest possible range of the attribute. A local range refers to the range of an attribute in a particular experiment. We propose that local range effects might be greatest when the global range is not well defined or does not exist. In Experiment 1, subjects rated the attractiveness of apartments on the basis of monthly rent and distance to campus. Neither rent nor distance has well-established global ranges, although some ranges are certainly more reasonable than others. In Experiment 2, subjects rated the overall performance of students on the basis of percentage scores for two exams. The global range of percentage scores is well defined; it goes from 0% to 100%. In Experiment 3, subjects rated the desirability of applicants for research assistantships on the basis of Scholastic Aptitude Test (SAT) scores and a motivation test. The global range of SAT is well defined; it goes from 200 to 800. Furthermore, the global distribution of SAT is well defined;
it is normally distributed. Any given score contains information about both absolute and relative performance. We expected that local range effects would be most pronounced in Experiment 1, less pronounced in Experiment 2, and least pronounced in Experiment 3.

Experiment 1: Apartments

Method

Subjects rated the attractiveness of apartments described by monthly rent, distance to campus, and the opinion of a friend who saw the apartment. There were three range conditions, with different subjects serving in each. The ranges of rent and distance varied across conditions, with a common set of apartments in each condition. In one condition, the ranges of rent and distance were both narrow. In a second condition, the range of rent was narrow, and the range of distance was wide. In a third condition, the range of rent was wide, and the range of distance was narrow. The range of the friend's opinion was held constant in all three conditions.

Stimuli and design. Apartments were described by either one attribute, two attributes, or three attributes. This design provides one method to test between additive and averaging processes (Birnbaum, 1980; Norman, 1976). Levels of rent in the narrow range were $200, $250, $300, $350, and $400. Levels of distance in the narrow range were 10, 14, 18, 22, and 26 min of travel time to campus. When the range of rent was wide, four additional levels ($100, $600, $800, and $1,000) were added to the narrow set. When the range of distance was wide, four additional levels (1, 30, 40, and 50 min) were added to the narrow set. Values of the friend's opinion were "Poor," "Average," and "Excellent."

A common set of 143 apartments (described by either one, two, or three attributes) appeared in each condition. Levels of rent and distance common to all three conditions corresponded to those in the narrow range. Thus, in the narrow ranges, there were no additional apartments beyond the common set. In the wide ranges, there were 96 additional apartments used to extend the ranges of the rent and distance attributes.

Instructions. Subjects made their responses on a category rating scale from 1 to 50, where 1 was not at all attractive and 50 was extremely attractive. They were instructed to use any integers between 1 and 50, inclusive. Subjects were told the endpoints for rent and distance and the three levels of opinion. In addition, subjects were informed that when apartments were described by fewer than three attributes, the missing information was unavailable; they were asked to make their ratings only on the basis of the information provided. Furthermore, subjects were told that the distance information referred to their current means of transportation. No other method of travel was available, and no shorter route to campus was possible.

Procedure. Subjects performed the experiment on IBM PC-XT computers. First, they were given instructions that explained the task and presented the endpoints of each attribute. Ten practice trials were presented, followed by a set of experimental trials. After completing the experimental trials, subjects allocated 100 points among the three attributes to reflect the relative importance of each attribute in their previous judgments. Subjects then wrote a brief paragraph describing how they did the task.

Participants. In all three experiments, subjects were undergraduates at the University of California at Berkeley, who received credit in psychology courses for their participation. There were 29 different subjects in each of the three range conditions. (A few additional subjects were not included in the analyses due to computer malfunction or failure to follow instructions.)

Results and Discussion

Range effects. Figure 6 presents range effects for single and multiattribute judgments. Mean attractiveness ratings for the common apartments are shown with a separate curve for each range. Results for distance can be seen in Panels A and B, and results for rent are shown in Panels C and D. To simplify, results are only presented for two conditions—the narrow rent and wide distance condition and the wide rent and narrow distance condition. Data are represented with solid points and solid lines; dashed lines depict predictions of a theory that will be discussed later.
For single-attribute judgments (Panels A and C), the effects of distance and rent are greater in the narrow range than in the wide range. The Attribute Level × Attribute Range interaction is statistically significant for rent, \( F(4, 224) = 4.46 \), but not for distance, although the pattern of results is similar for both attributes.\(^3\) Range effects in multiattribute judgments (Panels B and D) resemble those found with single-attribute judgments. Attribute Level × Attribute Range interactions were statistically significant in both panels: for distance, \( F(4, 224) = 4.61 \), and for rent, \( F(4, 224) = 2.76 \).\(^4\)

**Averaging effects.** There is another trend that can be seen in Figure 6. A comparison of single-attribute and multiattribute curves from the same context shows evidence of averaging. Averaging implies that the effect of an attribute is greater when presented alone (single-attribute judgments) than when presented with other information (multiattribute judgments). This prediction is supported in the data. For example, in the narrow context, the slope of the single-attribute curve for rent (Panel A) is steeper than the slope of the multiattribute curve for rent (Panel B). Similar effects can be seen with the wide context. Of the 24 possible tests of averaging that are based on comparisons of slopes, 19 were consistent with averaging. These comparisons support the claim that subjects average the information to arrive at overall evaluations.

Averaging effects cannot be described by additive models. Additivity implies that the effect of an attribute is independent of the other attributes with which it is presented. The effect of an attribute presented alone should be identical to the effect of an attribute presented with other information; slopes of the single-attribute curves should be the same as slopes of the multiattribute curves. Results in Figure 6 are inconsistent with an additive process, but the data can be described by an averaging process.

**Trade-offs between attributes.** To investigate the locus of range effects, it is necessary to examine the rank orders of judgments assigned to the same stimulus combinations in different ranges. If rank orders vary between contexts, range effects could not be described by changes in the judgment function alone.

Mean attractiveness ratings for the same set of apartments were compared in two range contexts. Figure 7 shows rank orders of means for the rent by distance design in the wide rent and narrow distance context (Panel A) and the narrow rent and wide distance context (Panel B). Arrows represent attractiveness orders of adjacent apartments for which one apartment has a lower rent and the other has a shorter distance. Arrows point toward the less attractive apartment. Downward arrows imply that cheaper apartments are ranked higher than closer apartments; upward arrows indicate that closer apartments are ranked higher than cheaper apartments.

Figure 7 shows numerous rank order reversals. When the range of rent is narrow (Panel A), arrows tend to point down. People tend to favor cheaper apartments over closer apartments. When the range of rent is wide (Panel B), arrows point up. People prefer closer apartments to cheaper ones. A given difference along an attribute (either rent or distance) has a greater effect in the narrow range than in the wide range. These effects produce changing trade-offs across contexts.

Many other rank order shifts occur besides those illustrated with arrows. For example, in Panel A, a $200 apartment 26 min from campus is judged more attractive than a $400 apartment 10 min away. The rank orders are 17 and 10. However, in Panel B, the opposite order occurs. The $200 apartment that is 26 min away is judged less attractive than the $400 apartment that is 10 min away. Rank orders are 9 and 20.

**Weights versus scales.** Differing trade-offs could be due to changes in either weights or scales. Weights can be distinguished from scales in an averaging model. Because there is considerable evidence of averaging, we use the procedure developed for the relative-weight averaging model to diagnose the locus of the effects.\(^5\) First, it is necessary to demonstrate a Rent × Rent Range interaction and a Distance × Distance Range interaction, as shown in Figure 6. Then one can test the Rent Range × Opinion interaction and the Distance Range × Opinion interaction.

If attribute range influences weight, there should be a significant interaction between attribute range and opinion for both rent and distance. If attribute range influences scales but not weights, there should be no interaction between attribute range and opinion for either rent or distance.

---

\(^3\) All significance tests were performed with a .01 level of alpha.

\(^4\) Significant effects were also found in the other subdesigns. For example, in the rent by distance designs, there were significant Attribute Level × Attribute Range interactions, \( F(8, 336) = 2.99 \) for distance and \( F(8, 336) = 4.21 \) for rent.

\(^5\) If the \( J \) function is assumed to be linear, the relative-weight averaging model predicts no interactions between attributes in the multiattribute designs. The three-way Rent × Distance × Opinion interaction was nonsignificant, but two of the two-way interactions were significant: for Opinion × Distance, \( F(8, 448) = 11.70 \), and for Opinion × Rent, \( F(8, 448) = 2.94 \). In the designs based on two attributes, two interactions were significant: Opinion × Distance, \( F(8, 448) = 3.09 \), and for Opinion × Rent, \( F(8, 448) = 2.71 \). However, the proportion of total variance in each interaction was less than 0.5%, so we will use the procedure for relative-weight averaging despite the interactions.
Attribute Range × Opinion interactions are nonsignificant for both attributes. In summary, results are consistent with the hypothesis that range effects can be described by changes in the subjective values of the attributes.

Judged weights. Table 1 shows mean judged weights for rent, distance, and opinion in the two range conditions. Average weights for rent and distance do not differ significantly in either range context. Although clearly not diagnostic of a scale change, the nonsignificant differences between judged weights in different range contexts are consistent with the interpretation that scales, rather than weights, vary with attribute range.

Model testing. The relative-weight averaging model (Equation 1) was fit to mean ratings of the common stimuli from both range conditions with the help of Chandler's (1969) STEPIT subroutine. For each condition, the loss index, \( P_c \), was defined as follows:

\[
P_c = \frac{\sum (R - \hat{R})^2}{\sum (R - \bar{R})^2},
\]

where \( P_c \) represents the proportion of residual variance for Condition \( c \), \( R \) is the mean response, \( \hat{R} \) is the prediction from Equation 1, \( \bar{R} \) is the grand mean for a condition, and the summation is over the 143 common trials in both conditions. A lack-of-fit index was computed for each condition, and the average value of the two indices, \( P_c \), was minimized.

Range effects were attributed to scale values; different scales were estimated for narrow and wide contexts. Furthermore, \( w_i \) was fixed to 1.0, and all other weights were estimated. When scales were allowed to vary with attribute range, \( P_c \) was 3.5% with 27 estimated parameters. For purposes of comparison, another version of the model was fit to the data that allowed both weights and scales to vary with attribute range. For this model, \( P_c \) was still 3.5%, with 29 parameters. In short, there was no improvement in fit when weights were also allowed to vary.

Predictions of the model that allows scales to vary with attribute range are presented as dashed lines in Figure 6. The model can account for range effects in single and multiattribute judgments. It can also describe the changing trade-offs across range contexts in Figure 7. Predicted rank orders (not shown) are similar to the observed ranks in both conditions.

In conclusion, similar range effects are found in both single and multiattribute judgments; a given difference along an attribute has a greater effect in the narrow range than in the wide range. Furthermore, range effects are not simply metric; they are also ordinal. Rank orders of the same stimuli differ between range contexts. Results are consistent with the hypothesis that scales depend on attribute range; neither judged weights nor estimated weights varied greatly. Experiment 2 examined range effects with attributes that have more clearly defined global ranges.

Experiment 2: Class Performance

Method

Subjects evaluated the overall class performance of students described by percentage correct scores on two exams (Exams A and B) and an essay grade. Each subject participated in one of two range conditions. In one condition, the range of Exam A was wide, and the range of Exam B was narrow. In the other condition, exam ranges were reversed. The range of essay grades was held constant in both conditions.

Stimuli and design. Students were described by either one attribute, two attributes, or three attributes, as described in Experiment 1. When the range of an exam was narrow, levels were 35%, 45%, 50%, 55%, and 65%. These levels also defined the range of common levels for both exams in each condition. When the range of an exam was wide, four additional levels (10%, 20%, 80%, and 90%) were included with the narrow set. Essay grades were A, C, and F. There were 239 experimental trials in each condition.

Instructions and procedure. Evaluations were made on a 1 to 9 category rating scale, where 1 was "Extremely Poor" and 9 was "Extremely Good." Subjects were told that the exams covered material of equal difficulty and had the same number of items. Subjects were also told that the exams were graded differently, and therefore, the range of scores differed on the two exams. The range of scores for each exam was provided. Otherwise, the procedure was identical to that used in Experiment 1.

Participants. There were 30 different subjects in each condition.

Results and Discussion

Range effects. Figure 8 shows range effects in single-attribute judgments (Panels A and C) and multiattribute judgments (Panels B and D). Mean performance ratings are shown for the common students with a separate curve for each range. Data are represented with solid points and solid lines; dashed lines show predictions of the changing-scale hypothesis. For single-attribute judgments, Exam A and Exam B have a greater effect in the narrow range than in the wide range. The Attribute Level × Attribute Range interaction was statistically significant for Exam A and Exam B, \( F(4, 232) = 12.29 \) and \( F(4, 232) = 29.85 \), respectively.

Range effects in multiattribute judgments resemble those found in single-attribute judgments. Interactions between attribute levels and attribute ranges are statistically significant in both panels, \( F(4, 232) = 8.82 \), and \( F(4, 232) = 15.85 \). Range effects in Experiment 2 resemble those in Experiment 1. Even when attributes have a well-established global range, a given attribute difference has a greater effect in the narrow range than the wide range.

Averaging effects. A comparison of single and multiattribute curves from the same context shows evidence of averaging. If subjects average, the effect of any given attribute is greater when presented alone than when presented with additional information. Averaging implies that for a given range context, the slope of the single-attribute curve should be steeper than the slope of the multiattribute curve.

<table>
<thead>
<tr>
<th>Context</th>
<th>Rent</th>
<th>Distance</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow, wide</td>
<td>32.6</td>
<td>41.2</td>
<td>26.2</td>
</tr>
<tr>
<td>Wide, narrow</td>
<td>39.1</td>
<td>33.8</td>
<td>27.1</td>
</tr>
</tbody>
</table>
This prediction is satisfied in the data. For example, in the narrow context, the single-attribute curve for Exam A (Panel A) is steeper than the multiattribute curve for Exam A (Panel B). Similar results can be observed in the wide context. Of the 24 possible tests of averaging that are based on comparisons of slopes, all 24 were consistent with averaging. These comparisons support the notion that subjects average information when arriving at overall performance ratings.

Trade-offs between attributes. Performance judgments of the same students were compared between range contexts. Figure 9 shows the rank orders of mean performance ratings when Exam A has a narrow range and Exam B has a wide range (Panel A) and when Exam A has a wide range and Exam B has a narrow range (Panel B). Arrows show orderings for adjacent students with trade-offs between exams. In Panel A, 10 out of 16 arrows point down; a given change in Exam A is favored over a given change in Exam B. In Panel B, two arrows point down; a change in Exam B is favored over a change in Exam A. There are numerous rank order shifts between range contexts.

Weights versus scales. Differing trade-offs across range contexts could be attributed to changes in weights or scales. If weights vary with range, there should be an interaction between essay and Exam A range and an interaction between essay and Exam B range. The curves should be inversely related to those in Figure 8. If scales vary with attribute range, there should be no interaction. The Essay \times Exam Range interaction is nonsignificant for both Exam A and Exam B. Attribute range appeared to influence the scale values, as found in Experiment 1.6

Judged weights. Table 2 shows mean judged weights for Exam A, Exam B, and essay in the two range contexts. Neither pair of judged weights differed significantly across range contexts. These results seem generally consistent with the change-of-scale interpretation.

Model testing. Mean performance ratings from both conditions were fitted to the relative-weight averaging model. Scales were allowed to vary linearly with attribute range. With this restricted scale-change model, \( P_6 \) was 3.1%, with 21 parameters. When both weights and scales varied, \( P_6 \) was 2.6%, with 23 parameters. The fit improved slightly, but not much, when weights, as well as scales, were allowed to vary with attribute range.

Predictions of the model that allows scales to vary with attribute range are presented as dashed lines in Figure 8. The model accounts for the steeper slopes of the curves in the narrow range relative to the wide range. It can also describe the changing rank orders in Figure 9. Predicted ranks were similar to observed rank orders.

In summary, the results of Experiment 2 showed that attribute range influences multiattribute judgments even when the global range is well defined. Range effects were similar to those found in Experiment 1, where the global range was more ambiguous. In Experiment 3 we examined range effects with an attribute that not only has a well-defined global range, but also a well-defined global distribution—namely, SAT scores.

---

6 We are assuming that the relative-weight averaging model describes the data when testing between weight and scale hypotheses. When the \( J \) function is assumed to be linear, the relative-weight averaging model implies that there should be no interactions between attributes in the multiattribute designs. Of the seven possible interactions, three were statistically significant: for Essay \times Exam B, \( F(8, 464) = 4.28 \); for Essay \times Exam B in the three-attribute design, \( F(8, 464) = 3.73 \); and for Exam A \times Exam B in the three-attribute design, \( F(16, 928) = 2.29 \). Once again, the proportion of total variance was less than 1% in each interaction, so the relative-weight averaging model was assumed.
Experiment 3: Research Assistantships

Method

Subjects rated the desirability of undergraduate applicants for research assistantships. Applicants were described by a verbal SAT score, a score on a motivation test, and the number of semesters of previous research experience. There were two range conditions. In one condition, the range of SAT was wide and the range of motivation was narrow. In the other condition, ranges were reversed. The range of experience was the same in both conditions.

Stimuli and design. Applicants were described by either one attribute, two attributes, or three attributes. When the range of SAT scores was narrow, levels were 400, 450, 500, 550, and 600. When the range was wide, four additional levels (200, 300, 700, and 800) were included. When the range of motivation scores was narrow, levels were 20, 25, 30, 35, and 40. When the range was wide, four additional values of 0, 10, 50, and 60 were included. Levels of both SAT and motivation scores common to the two range contexts were identical to levels used in the narrow ranges. Levels of previous experience were 1, 2, and 3 semesters.

Instructions and procedure. Subjects rated the performance of applicants on a category rating scale ranging from very poor (1) to very good (9). Instructions gave the endpoints for each attribute (200 to 800 for verbal SAT and 0 to 60 for the motivation test) and the range of scores of the applicants presented for judgment. Subjects were told that SAT scores were normally distributed, and the density function was presented graphically. Furthermore, subjects were informed that SAT scores had a mean of 425 and a standard deviation of 100, and that the motivation test measured the extent to which a student initiates activities and fulfills goals.

Participants. There were 35 different subjects in each condition.

Results and Discussion

Range effects. Figure 10 shows range effects in single-attribute judgments (Panels A and C) and multiattribute judgments (Panels B and D). A given difference along an attribute has a greater effect when presented in a narrow range than in a wide range. Interactions between attribute level and attribute range were statistically significant in all four panels: $F(4, 272) = 3.75$, $F(4, 272) = 3.84$, $F(4, 272) = 7.61$, and $F(4, 272) = 8.12$, respectively. Even when an attribute has a well-defined global distribution, range effects occur, although the magnitude of the effects for SAT seems to be somewhat smaller than that found with other attributes.

Averaging effects. Figure 10 also shows evidence of averaging. For example, in the narrow range context, the slope of SAT when presented alone is steeper than the slope of SAT when presented with other attributes. Of the 24 possible tests of averaging that are based on comparisons of slopes, 22 were consistent with an averaging hypothesis.

Trade-offs between attributes. Judgments of the same applicants were compared between range contexts. Figure 11 shows mean desirability ratings for the SAT by motivation design in the two ranges. Arrows point toward the applicant judged less desirable. In Panel A, 5 arrows point up, and in Panel B, 9 arrows point up. There are some rank order shifts between ranges, although fewer than in the previous two experiments.

Weights versus scales. To determine whether range effects could be attributed to weights or scales, the Experience X SAT Range interaction and the Experience X Motivation Range interaction were examined. Only 1 of the 7 possible interactions (SAT X Motivation) was statistically significant, $F(16, 1,088) = 2.15$, so the relative-weight averaging model with a linear $J$ was assumed.
were 35.6 and 41.5. Neither pair of judged weights differed significantly, consistent with the changing scale interpretation.

Model testing. Mean desirability ratings from both conditions were fitted to the relative-weight averaging model. When scales were allowed to vary linearly with attribute range, $P_c$ was 2.7%, with 21 parameters. When both weights and scales were allowed to vary, $P_c$ was 2.4%, using 23 parameters. Results were consistent with the changing-scale hypothesis; neither estimated nor judged weights varied greatly with attribute range. Predictions of the averaging model that allowed changing scales are presented as dashed lines in Figure 10. In summary, Experiment 3 showed that range effects occur even when an attribute has a well-defined global distribution. Once again, range effects were attributed to changes in scales.

General Discussion

Many types of judgments, from clinical assessments to investment decisions, require the decision maker to consider multiple attributes. In the present article, we investigated whether attribute range influences multiattribute judgments. Range effects were demonstrated with both single and multiattribute responses along a variety of different dimensions. A given difference along an attribute had a greater effect in a narrow range than in a wide range. Furthermore, rank orders of judgments assigned to the same stimulus combinations varied across ranges. Trade-offs between time and money, one measure of ability and another, and achievement and motivation can differ, depending on attribute ranges.

Weight Invariance

Past research on range effects in multiattribute judgment has tended to focus on the extent to which weights vary with attribute range (Beattie & Baron, 1991; Fischer, 1991; Gabrielli & von Winterfeldt, 1978; Levin, Kim, & Corey, 1976; Weber, 1992). General conclusions, however, are hard to draw. Although most of these studies argue that weights do not vary with attribute range, range effects have not been demonstrated. Without demonstrating that range effects occur, one cannot determine whether weights are sensitive to attribute range. When range effects are nonsignificant, it is a foregone conclusion that weights and scales will be invariant. Furthermore, range effects could be nonsignificant for many reasons other than a true null effect. For example, past studies on range effects did not use identical attribute levels in different attribute ranges. Attribute levels were confounded with attribute ranges, and this confounding increases the error variance.

Another issue that complicates general conclusions from past research is that weight is defined differently across studies. Beattie and Baron (1991) defined weights as regression coefficients and argued that weights do not depend on attribute range. Others defined weights as judged importances. In several studies, researchers have demonstrated that judged weights are insensitive to changes in attribute range (Fischer, 1991; Gabrielli & von Winterfeldt, 1978; Stewart & Ely, 1984; von Nizszch & Weber, in press). These measures of importance have been questioned by some researchers because they do not correlate with regression weights in predictable ways. Nisbett and Wilson (1977) assessed people's understanding of their decisions by comparing regression weights with judged weights. The low correlation between these measures led them to conclude that people lack insight into their own mental processes.

Several researchers (Birnbaum, 1974a; Birnbaum & Stegner, 1981; Goldstein, 1990) pointed out problems with regression weights as measures of importance and discussed why those weights should not necessarily correlate with judged weights. Regression weights do not necessarily reflect the psychological importance of attributes because they vary with the unreliability of the response, the stimulus metric, the response metric, the covariance with other predictors, and other factors (Birnbaum, 1974b). Furthermore, in a regression analysis, weights and subjective values are confounded.

In the present article, we defined estimated weights as parameters of an averaging model (w in Equation 1). The averaging model provides an explicit psychological description of how information is combined, and it can easily be rejected by the data. Estimated weights from the averaging model correlated highly with judged weights. Birnbaum and Stegner (1981) demonstrated that judged weights correlated

![Table 3](Image)

<table>
<thead>
<tr>
<th>Context</th>
<th>SAT</th>
<th>Motivation</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow SAT, wide motivation</td>
<td>34.4</td>
<td>41.5</td>
<td>24.1</td>
</tr>
<tr>
<td>Wide SAT, narrow motivation</td>
<td>40.4</td>
<td>35.6</td>
<td>24.0</td>
</tr>
</tbody>
</table>

Note. SAT = Scholastic Aptitude Test.

8 A reanalysis of the data from Beattie and Baron (1991) shows range effects in Experiments 1, 4, and 6. Furthermore, these effects tend to be in the same direction as those obtained in the present article. A given attribute difference has a greater effect when presented in a narrow range than in a wide range.
strongly with estimated weights in judgments of the IQ of an adopted child, given information about biological parents’ and adopted parents’ IQs. Results from the present studies show that estimated weights from an averaging model, as well as judged weights, do not tend to vary greatly with attribute range.

Weight Changes

If attribute range does not influence estimated weights from the averaging model, what factors do? Weights from averaging models depend on the validity and reliability of the attributes (Birnbaum & Mellers, 1983). When subjects predict numerical criteria from numerical cues, weights vary systematically with the cue-criterion correlation (Birnbaum, 1976). When subjects judge the likeableness of a person on the basis of trait adjectives provided by sources, the weight of the trait adjective varies with the length of acquaintance of the source (Birnbaum et al., 1976). When subjects predict student exam performance from IQ tests and study time, weights vary with the reliability of the IQ and study time measures (Surber, 1981).

Tversky et al. (1988) argued that weights, defined by regression coefficients, vary systematically with the compatibility of the attribute and the response. In a study by Slovic, Griffin, and Tversky (1990), subjects were asked to predict the performance of students in a course on the basis of their grades in another course and their class ranks in a third course. Half of the respondents predicted grades, and the other half predicted class ranks. The compatibility hypothesis asserts that the regression weight for grades should be larger than the regression weight for class ranks when subjects are predicting grades, and the opposite order of regression weights should occur when subjects are predicting class ranks. Slovic et al. (1990) presented results consistent with this hypothesis.

An alternative interpretation of the data is that more valid predictors receive greater weight. Grades are more valid predictors of grades, and class ranks are more valid predictors of class ranks. This interpretation makes no claims about compatibility effects per se. In a recent article, Cooke and Mellers (1994) separated compatibility effects from range effects and found that attribute range influenced multiattribute judgment, as demonstrated in the present article, but compatibility effects were nonsignificant. Thus, compatibility effects were negligible when unconfounded with attribute range. It may also be the case that compatibility effects are negligible when unconfounded with other factors, such as the reliability and validity of the attributes.

Scale Changes

What factors influence scale values? Scale values vary with contextual effects due to stimulus spacing (Mellers & Birnbaum, 1982, 1983). Stimulus spacing refers to the distribution of levels along an attribute. When the distribution is positively skewed, scale values for common levels can differ from those obtained when the distribution is negatively skewed. When subjects combine stimuli along different dimensions (e.g., salaries and merit indexes of faculty members), rank orders of judgments differ between contexts, and scales vary. Mellers and Birnbaum assumed that, in these cases, contextual effects occur before stimulus combination (H in Figure 1).

When stimuli along the same dimension are combined or compared, scale values do not appear to vary with the stimulus spacing. Rank orders of “ratio” and “difference” judgments of the darkness of dot patterns were similar between contexts having different marginal dot distributions. In these cases, contextual effects occur after stimulus combination (J in Figure 1).

Birnbaum and Stegner (1979) demonstrated that scale values can also depend on the bias of the source providing the information. Suppose a subject is asked to rate the desirability of nuclear power guidelines. Information is provided by a Sierra Club president and an engineer representing the local utilities company. Subjects seemed to adjust the subjective value of the information up or down, depending on the bias of the source.

Implications

Rational theories of consumer choice imply that indifference curves should not cross. A preference ordering can be derived over bundles of goods when four assumptions are satisfied: more of a good is better than less, goods have diminishing marginal utility, preferences are transitive, and preferences are complete (Frank, 1991). Taken together, these assumptions rule out crossing indifference curves. The present results show that indifference curves that are based on multiattribute judgments can cross, depending on range contexts. The rate at which subjects are willing to substitute one attribute for another varies with attribute range. In short, preference orderings for the same stimuli can change across contexts that differ only in attribute range.

Knowledge that trade-offs can change with attribute range may be useful in marketing applications. If a company introduces a new product that is higher than the competitor in both quality and price, there are two ways to market the new product. One strategy enhances perceived differences in quality. Marketing slogans such as “Taste the difference!” and “Quality is Job 1” may serve this purpose. The other strategy minimizes perceived differences in price. A commercial for a telecommunications company stacks up a small set of pennies, and a voice announces, “This is all you save with other companies. Isn’t the difference worth it?”

The present results demonstrate that rank orderings of the same stimuli can be reversed if appropriate attribute differences are highlighted.

Finally, those who use multiattribute utility theory in applied domains should consider the implications of attribute range on preference. If a decision maker is comparing potential locations for a water treatment plant, the preference order for two locations can be altered depending on which attribute range is selected (von Winterfeldt & Edwards, 1986). Furthermore, if new locations enter the
choice set after the utilities have been elicited, it may be unwise to assume that the same utilities apply. The present results show that utilities can vary with attribute range.

Conclusions

This article demonstrates range effects in both single and multiattribute judgments. Local range effects were found with attributes having poorly defined global ranges, such as rent, and well-defined global ranges, such as percentage scores. In all of the experiments, rank orders of the same stimulus combinations varied across range contexts. Stretching and shrinking the range altered the trade-offs between attributes.

Some have argued that changes in weights can lead to rank order shifts, and rank order shifts imply weight changes. Whereas it is true that changes in weights can lead to rank order reversals, the present experiment shows that changes in scales can also produce rank order reversals. The locus of range effects is in the subjective values of the attribute, and neither estimated nor judged weights vary greatly with attribute range. Scales tend to vary linearly with attribute range, consistent with a range–frequency view of the mapping from physical value to psychological value.

References


Received December 14, 1992
Revision received November 1, 1993
Accepted February 16, 1994

---

### 1995 APA Convention Call for Programs

The *Call for Programs* for the 1995 APA annual convention appears in the September issue of the *APA Monitor*. The 1995 convention will be held in New York, New York, from August 11 through August 15. The deadline for submission of program and presentation proposals is December 2, 1994. Additional copies of the *Call for Programs* are available from the APA Convention Office, effective in September. As a reminder, agreement to participate in the APA convention is now presumed to convey permission for the presentation to be audiotaped if selected for taping. Any speaker or participant who does not wish his or her presentation to be audiotaped must notify the person submitting the program either at the time the invitation is extended or before the December 2 deadline for proposal submission.