

Testing Interval Independence Versus Configural Weighting Using Judgments of Strength of Preference

Michael H. Birnbaum
California State University, Fullerton, and
Institute for Mathematical Behavioral Sciences

Laura A. Thompson and David J. Bean
California State University, Fullerton

Judges made choices and rated strengths of preference between gambles composed of 50–50 chances to receive either of 2 monetary outcomes (x, y). Others judged how much they would pay to play their chosen gamble rather than the other gamble. Judged strengths of preference violated interval independence, because they depended on the value of a common outcome. For example, judges offered to pay an average of \$44 to play (\$74, \$100) instead of (\$8, \$100) but offered to pay only \$24 to play (\$6, \$74) instead of (\$6, \$8). Results violated the theory that utility of gambles is a nonconfigural average of the values of the outcomes and that strengths of preference are monotonically related to utility differences. Results can be explained by a configural weight model in which the lowest outcome receives greater weight.

Most people think 50–50 gambles are worth less than their expected values. For example, many people would prefer a sure \$40 rather than accept a 50–50 gamble that would pay either \$0 or \$96, even though the expected value of the gamble is \$48. It is common for a person to offer to pay only \$24 for a 50–50 chance to win either \$0 or \$96 (e.g., Birnbaum & Sutton, 1992). Because people offer less and will accept less than expected value for a gamble involving risk, they are described as “risk averse.”

According to classical utility theory (von Neumann & Morgenstern, 1947), such risk aversion occurs because the utility of money is a nonlinear function of money, specifically concave downward; therefore, the utility of the expected value exceeds the expected utility of the gamble. For example, the expected utility of the (\$0, \$96) gamble (i.e., the average value of utility of the monetary outcomes) is $U(\$0, \$96) = [u(\$0) + u(\$96)]/2$. Suppose the utility of money is the square root of money, $u(x) = x^{.5}$. Then the expected utility of the gamble is $U(\$0, \$96) = 4.9$; applying the inverse function to both sides, the cash equivalent of the gamble would be $(4.9)^2 = \$24$. In comparison, the expected

value is \$48 $[(\$0 + \$96)/2]$, which has a higher utility (6.93) than the expected utility of the gamble (4.9). Because the utility of the expected value exceeds the expected utility, this theory can account for risk aversion (preferring the expected value of a gamble to the gamble itself).

However, configural weight theory can explain risk aversion without a nonlinear utility function (Birnbaum, Coffey, Mellers, & Weiss, 1992; Birnbaum & Sutton, 1992; Wakker, 1994). Configural weight theory, along with rank and rank-and-sign-dependent utility theories, allows the weight of a stimulus component to depend on the rank of that stimulus component among the other components that comprise a stimulus (Birnbaum, 1974; Birnbaum et al., 1992; Lopes, 1990; Luce, 1992; Luce & Fishburn, 1991; Luce & Narens, 1985; Tversky & Kahneman, 1992; Wakker, 1993; Wakker, Erev, & Weber, 1994; Weber, 1994).

To illustrate how risk aversion can be explained with a linear utility function, suppose $u(x) = x$. Suppose also that the weights of the lower and higher outcomes in a 50–50 gamble are .75 and .25, respectively, instead of .5 and .5 (as they would be in expected utility theory). Then configural weight theory predicts that the utility of the gamble is $U(\$0, \$96) = .75u(0) + .25u(96) = .25(96) = 24$, [which also has a cash equivalent of \$24, since $u(x) = x$]. In this example, the weight of \$96 was only .25 because it was the higher outcome; however, if the gamble were a 50–50 gamble between \$96 and \$192, then \$96 would be the lower valued outcome, and \$96 would receive a weight of .75 instead of .25. In this configural weight theory, the weight of an outcome is independent of its value per se but depends on its rank in comparison with the other outcomes of the same gamble.

Thus, both configural and nonconfigural theories can account for risk aversion, but they do so using very different utility functions. Unless the utility function is known or constrained, it can be difficult to distinguish configural weight theory from utility theory (Birnbaum et al., 1992);

Michael H. Birnbaum, Department of Psychology, California State University, Fullerton, and Institute for Mathematical Behavioral Sciences; Laura A. Thompson and David J. Bean, Department of Psychology, California State University, Fullerton.

Laura A. Thompson is now a graduate student at Baylor University.

Thanks are due to Mary Kay Stevenson and Peter Wakker for helpful comments on a draft. Support was received from National Science Foundation Grant SES 8921880. Completion of the manuscript was facilitated by National Science Foundation SBR-9410572.

Correspondence concerning this article should be addressed to Michael H. Birnbaum, Department of Psychology, California State University, Fullerton, P.O. Box 6846, Fullerton, California 92834-6846. Electronic mail may be sent via Internet to mbirnbaum@fullerton.edu.

therefore, it can be difficult to determine the utility function in a more general model that allows both weights and utility functions to be estimated from the data (Birnbbaum & Sutton, 1992). The present article extends a technique developed in an analogous situation to distinguish configural from nonconfigural theories in another judgment domain (Birnbbaum, 1974). The technique uses judgments of strengths of preference between alternatives to test interval independence.

In this study, two methods are used to elicit judgments of strength of preference between gambles. One method is to ask people to rate degrees of preference directly, using a category rating scale, as in Birnbbaum (1974). Another method is to ask people how much they would pay to receive one gamble rather than another. When people will pay money to receive *A* instead of *B*, it seems reasonable to theorize that they prefer *A* to *B*. It also seems reasonable to theorize that the more people will pay, the greater is the difference in utility between alternatives *A* and *B*. The following equation expresses these assumptions:

$$D(A; B) = J[U(A) - U(B)], \tag{1}$$

where $D(A; B)$ is the amount that a person would pay to get *A* instead of *B*; $U(A)$ and $U(B)$ are the psychological utilities of *A* and *B*; and J is a strictly monotonic function. If the utility of money is a nonlinear function of money, then J would reflect this source of nonlinearity. Another source of nonlinearity in the J function might be inertia in willingness to pay. For example, Equation 1 allows a person to refuse to pay for a small improvement in utility, since J could be very flat when the difference is near zero. An analogous equation can be written for ratings of preference, which would be expected to involve a different J function, mapping differences in utility into ratings of preference. Previous research has found that the J function can be S-shaped when the preference response is on a bounded scale (Rose & Birnbbaum, 1975; Stevenson, 1986).

Additive Models Imply Interval Independence

Using Equation 1 as a premise, it is possible to get a powerful test of nonconfigural versus configural theories of the utility of risky gambles. This test uses Birnbbaum's "scale-free" method, which is more fully described in Birnbbaum (1974), Birnbbaum and Veit (1974), and Birnbbaum (1982). For the present case, consider a gamble between two equally likely monetary outcomes, x and y . A class of nonconfigural utility theories, including expected utility (EU), Savage's (1954) subjective expected utility (SEU) theory, and the psychological version of SEU (Edwards, 1954) can be written as follows:

$$U(x, z) = cu(x) + du(z), \tag{2}$$

where $U(x, z)$ is the (nonconfigural) weighted utility of the gamble to receive either x or z ; $u(x)$ and $u(z)$ are the utilities of monetary outcomes, x and z , respectively; c and d are constants (in the additive weighted utility theory if the two events are equally likely, $c = d$; nonadditive SEU would

allow $c + d < 1$; for additive SEU or EU theory, $c = d = 1/2$).

Now suppose that the alternatives, *A* and *B*, in Equation 1 are themselves gambles composed of two equally likely outcomes, such that $A = (x, w)$ and $B = (y, z)$. Suppose people are asked to compare gambles *A* and *B* and to judge how much they would pay to receive gamble *A* rather than gamble *B*. We can compose Equations 1 and 2 as follows:

$$\begin{aligned} D(A; B) &= J[U(x, w) - U(y, z)] \\ &= J[cu(x) + du(w) - cu(y) - du(z)] \\ &= J[c\{u(x) - u(y)\} + d\{u(w) - u(z)\}]. \end{aligned} \tag{3}$$

Equation 3 shows that the judgment of strength of preference between gamble *A* over *B* should be a function of the sum of two weighted differences in utility. If two gambles have a common consequence, $w = z$, then $u(w) - u(z) = 0$, so the degree of preference should be independent of that common outcome. Equation 3 becomes

$$D(x, z; y, z) = J[c\{u(x) - u(y)\}], \tag{4}$$

which is independent of the common outcome, z . The implication of equality of strengths of preference in Equation 4 is termed "interval independence." Interval independence would also be observed if people were to edit comparisons between gambles by eliminating common components (Tversky, 1972; Kahneman & Tversky, 1979).

Predicted Violation of Interval Independence

However, configural weight theories, including rank-dependent utility theories, do not require interval independence. Unlike EU and SEU theories, weights associated with the components depend on the relative utilities of the outcomes within the gambles. According to the configural weight model of Birnbbaum et al. (1992), lower valued outcomes typically receive greater weight than the higher of two positive outcomes. Birnbbaum et al. (1992) found that the lower outcome in a 50-50 gamble between two positive outcomes receives a weight of .73, .63, or .53 from the buyer's, neutral's, or seller's points of view, respectively (see also Birnbbaum & Sutton, 1992). This model predicts that when the common outcome is the lowest outcome in the set, the difference due to other contrasts will be less than when the common outcome is the highest outcome in the set.

According to rank-dependent theories (e.g., Birnbbaum, 1974; Birnbbaum et al., 1992; Birnbbaum & Stegner, 1979; Luce, 1992; Weber, 1994), Equation 2 can be rewritten as follows when $x > z$;

$$U(x, z) = au(x) + bu(z). \tag{5a}$$

However, when $x \leq z$,

$$U(x, z) = bu(x) + au(z). \tag{5b}$$

Therefore, composing Equations 1 and 5, we have two equations for strength of preference, depending on whether or not the common alternative is the lowest or highest

outcome of the gambles. When the common outcome is the lowest, $x > w$ and $y > z$ and $w = z$, the equation is as follows:

$$\begin{aligned} D(x, w; y, z) &= J[U(x, w) - U(y, z)] \\ &= J[au(x) + bu(w) - au(y) - bu(z)] \\ &= J[a\{u(x) - u(y)\}]. \end{aligned} \tag{6a}$$

However, when the common outcome is highest, $x < w$ and $y < z$ and $w = z$,

$$\begin{aligned} D(x, w; y, z) &= J[U(x, w) - U(y, z)] \\ &= J[bu(x) + au(w) - bu(y) - au(z)] \\ &= J[b\{u(x) - u(y)\}]. \end{aligned} \tag{6b}$$

When the common value is the lower outcome in one gamble and the higher in the other, $x > z = w > y$, the difference will be given as follows:

$$\begin{aligned} D(x, w; y, z) &= J[bu(x) + au(z) - bu(z) - au(y)] \\ &= J[b\{u(x) - u(z)\} + a\{u(z) - u(y)\}]. \end{aligned} \tag{6c}$$

Note that Equations 6a, 6b, and 6c are different, depending on the relative magnitudes of x and y and depending on a and b . If $a = b$, then Equation 4 follows as a special case. However, if the weight of the lower valued outcome within each gamble is greater than the weight of higher outcome ($b > a$), the strength of preference in Equation 6a (when the common outcome is lowest in value) will be less than the strength of preference in Equation 6b (when the common outcome is highest in value). Equation 6c is intermediate between 6a and 6b. (Note that if $b > a$, $J[b\{u(x) - u(y)\}] = J[b\{u(x) - u(z)\} + b\{u(z) - u(y)\}] > J[b\{u(x) - u(z)\} + a\{u(z) - u(y)\}] > J[a\{u(x) - u(z)\} + a\{u(z) - u(y)\}] = J[a\{u(x) - u(y)\}]$).

The purpose of this experiment is to test Equation 4 against Equations 6a, 6b, and 6c. Classical utility theory combined with Equation 1 implies that preference intervals should be independent of the common outcome, as in Equation 4. However, configural weight theory (assuming lower outcomes have higher weight) predicts that the difference in preference will be greater when the common value is the higher outcome than when it is the lower outcome.

Method

Instructions

The instructions read (in part) as follows:

This is an experiment on decision making. We are interested in how people choose between lotteries (chances to win or lose money). On each trial, you will be offered a comparison between two gambles. Your task is to decide which of the two gambles you would prefer to play and to judge how much you would pay to play your preferred gamble rather than the other gamble.

For example, suppose there is a can that contains exactly one \$10 bill and one \$20 bill . . . (you will get to select one

from the can blindly at random) . . . you have a fifty-fifty chance to receive either \$10 or \$20.

Now, would you rather reach in the can or would you prefer to reach in another can that contained one \$50 bill and one \$100 bill? This choice will be displayed as follows:

1. _____ (\$10, \$20) vs. (\$50, \$100)

In this case . . . you would prefer the second gamble (on the right). How much would you pay to play the gamble on the right rather than the gamble on the left? You should be willing to pay at least \$30 but no more than \$90 to play the gamble on the right . . . between these values, it is a matter of opinion.

. . . consider the following pair:

2. _____ (\$5, \$100) vs. (\$20, \$50)

. . . Some people would prefer the gamble on the left and some would prefer the gamble on the right. It is a matter of opinion. Circle the one you would prefer to play, then write down how much you would pay to play the preferred gamble rather than the other gamble in the space provided.

Rating Task Instructions

The rating task was run as a separate experiment, with different judges. The stimuli, design, and general procedures were the same, except that judges received the following instruction to define the response scale: "You will make your ratings on a scale from -9 to 9 to indicate the degree of your preference between the gamble on the left and the gamble on the right." The rating scale was labeled as follows: -9 = Prefer the gamble on the left very much more; -7 = Prefer the gamble on the left very much more; -5 = Prefer the gamble on the left much more; -3 = Prefer the gamble on the left more; -1 = Prefer the gamble on the left slightly more; 1 = Prefer the gamble on the right slightly more; 3 = Prefer the gamble on the right more; 5 = Prefer the gamble on the right much more; 7 = Prefer the gamble on the right very much more; 9 = Prefer the gamble on the right very very much more.

Stimuli and Design

The stimuli were displayed as in the examples above. The values were selected according to two designs, varying the values of x , y , w , and z . The first design used a common value, z , in comparisons of the following type: (x, z) versus (y, z) . The first design was a 3×3 , z by $(x$ and $y)$ factorial. The three levels of the common outcome, z , were either \$6, \$12, or \$100; and the values of y and x were either \$8 and \$74, \$8 and \$92, or \$10 and \$80, respectively.

The second design was included to break up the pattern of the first design and to allow checks on the consistency of the judgments. The second design consisted of all 21 comparisons of the following seven gambles: (\$12, \$96), (\$35, \$40), (\$24, \$84), (\$41, \$46), (\$36, \$72), (\$46, \$56), and (\$48, \$60). These seven gambles include four with equal expected value but different ranges, and they produce comparisons in which all four numbers differ, to help break up the pattern of the first design and presumably keep the judges attending to all four values.

Procedure

Trials from both designs were randomly intermixed with the restriction that any two trials from the first design were separated by at least one trial from the second design. These were printed with the instructions in booklets, along with a separate page of six

warm-up trials. The warm-up trials allowed the experimenter to check if the judges preferred dominant gambles. Some warm-up trials also allowed a check if judgments were in reasonable bounds. For example, the following comparison was included: (\$90, \$180) versus (\$35, \$80). The gamble on the left has both outcomes better than the outcomes on the right. The judge should be willing to pay at least \$10 but no more than \$145 to get the gamble on the left. People who preferred the gamble on the right or who gave a judgment outside this range were asked to reread the instructions and were asked to complete the warm-up trials again. When the warm-ups were completed, judges were instructed to complete the 30 experimental trials.

Judges

The judges were 157 undergraduates who received extra credit in introductory psychology for their participation. One hundred served in Experiment 1, in which strengths of preference were measured by the amount that the judges were willing to pay to receive the chosen gamble rather than the other gamble. There were 57 different undergraduates who served in the rating task of Experiment 2.

Results

Tests of Interval Independence Versus Configural Weighting

Table 1 shows mean judgments of the amounts judges were willing to pay to receive gamble (x, z) rather than gamble (y, z). If Equation 4 were correct, rows would be identical, except for error. Instead, mean judgments in the last row are consistently greater than corresponding values in the first two rows. The means in the last row are always at least \$20 more than the values in the other rows. The direction of this violation from Equation 4 is predicted by the configural weight model of Birnbaum et al. (1992), if lower outcomes receive greater weight.

The violations of interval independence (Equation 4) shown in Table 1 are also characteristic of the data of individuals. Of the 100 judges in Experiment 1, 86 had a larger mean for the last row (in which $z = \$100$) than for the first row (in which $z = \$8$); 83 of these judges also had their mean of the last row greater than the mean of the second row ($z = \$12$) as well. Analysis of variance (ANOVA) indicated that the main effect of rows in Table 1 is significant (in this article, "significant" refers to $p < .01$ throughout), $F(2, 198) = 117.5$. The main effect of columns

Table 1
Mean Judgment of Amount Willing to Pay to Play Gamble (x, z) Instead of (y, z)

Common value, z	Contrast (x vs. y)		
	\$74 vs. \$8	\$80 vs. \$10	\$92 vs. \$8
\$ 6	23.60	27.15	26.81
\$ 12	27.73	18.70	31.76
\$100	44.01	47.97	54.81

Table 2
Mean Rating of Strength of Preference for Gamble (x, z) Over (y, z)

Common value, z	Contrast (x vs. y)		
	\$74 vs. \$8	\$80 vs. \$10	\$92 vs. \$8
\$ 6	6.70	6.93	7.30
\$ 12	6.82	6.26	7.11
\$100	7.95	7.98	8.33

(x vs. y comparisons) was also statistically significant, $F(2, 198) = 17.9$.

Table 2 shows the comparable results for the rating task of Experiment 2. Again, the means in the last row exceed those of the first two rows. Of the 57 judges in this condition, 31 had a larger mean for the third row ($z = \$100$) than for the first row ($z = \$6$) and only 6 had the means in opposite order. ANOVA indicated a significant effect of rows, $F(2, 112) = 25.25$. Main effect of columns was also significant, $F(2, 112) = 8.25$. In both experiments, the values in the middle row are intermediate between the other rows, except for $x = \$90$ and $y = \$10$, producing a significant interaction in the first experiment, $F(4, 396) = 10.9$, but not in the second, $F(4, 224) = 1.74$.

In summary, both experiments show that the strength of preference is greater in the last row (where the common outcome is higher) than in the first row (where the common outcome is lower). These results are inconsistent with Equation 4 and consistent with Equations 6a, 6b, and 6c, if the weight of lower outcomes is greater, that is, if $b > a$.

Weak Transitivity and Scalability

Equation 1 also implies that if people will pay to get A rather than B , and if they will pay to get B rather than C , then they should be willing to pay to get A rather than C . This property, called weak transitivity, is defined as follows: If $D(A, B)$ and $D(B, C)$ are both positive, then $D(A, C)$ should also be positive. Equation 1 also implies a stronger form of transitivity, called scalability, which requires that if the above condition is met, then $D(A, C)$ should be at least as large as the larger of $D(B, C)$ and $D(A, B)$.

The mean judgments of strength of preference (paying prices) for the second design of Experiment 1 are shown in Table 3. The gambles have been listed in descending preference order in both rows and columns, and the signs have been adjusted so that a positive sign indicates preference for the column gamble over the row; negative signs would indicate that the row gamble is preferred to the column. After the gambles have been thus ordered, the mean payments are all positive, indicating that weak transitivity was satisfied by the mean judgments.

Strong transitivity (scalability) implies that magnitudes in each row or column should change monotonically. Instead, the means violate scalability. For example, a systematic deviation from scalability can be seen in Table 3, involving gambles G, F , and A . The means for $D(G, F)$, $D(F, A)$, and $D(G, A)$ are \$17.15, \$2.90, and \$6.58, respectively. Note

Table 3
Mean Judgment of Amount Offered to Receive Preferred Gamble (Design 2)

Gamble	Preferred gamble					
	G (48, 60)	E (36, 72)	C (24, 84)	F (46, 56)	A (12, 96)	D (41, 46)
E (36, 72)	6.8					
C (24, 84)	6.5	3.6				
F (46, 56)	17.2	13.2	3.3			
A (12, 96)	6.6	4.6	2.8	2.9		
D (41, 46)	19.6	14.2	9.2	13.2	7.2	
B (35, 40)	18.4	21.8	14.2	14.9	5.3	7.2

Note. Data based on 93 judges, who satisfied dominance in Design 1. Bold values show comparisons involving dominance.

that all three are positive, consistent with weak transitivity. However, because G (\$48, \$60) is preferred to F (\$46, \$56), which in turn is preferred to A (\$12, \$96), $D(G, A) = \$6.58$ should be at least as large as the biggest difference; instead, it is not as large as $D(G, F) = \$17.15$.

This pattern of violations of scalability may be attributable to the effects of transparent dominance. Because G (\$48, \$60) dominates F (\$46, \$56), it is an "easy" choice; however, the expected value of A (\$12, \$96) is greater than that of F , but F has a greater lowest outcome, so this choice is more "difficult." Six of the seven largest mean judgments in Table 3 involve dominant choices, pairs in which both outcomes are higher in one gamble than the corresponding outcomes in the other gamble. These cases often exceed expected value difference; for example, $D(G, F) = \$17.15$ exceeds the difference in expected values ($\$54 - \$51 = \$3$). Similar results were obtained for the ratings, with even greater exaggeration for pairs involving dominance (the modal and median ratings in these cases were all 9 or -9).

It is interesting to examine the (G, F, A) violation of scalability at the level of individuals. Out of 100 individuals in Experiment 1, 93 appeared to conform to dominance in Design 1; of the 93 remaining, 8 preferred F to G (violating dominance in Design 2), 8 had intransitive judgments as follows: $G > F > A > G$; 23 had a transitive order of $A > G > F$; and 11 had the transitive order $G > A > F$. Out of the 41 who had the transitive order $G > F > A$, 23 satisfied scalability for this triad, 12 had the pattern of violation of scalability shown by the means in Table 2, and 6 showed other violations. Thus, in this case, the pattern of the means appears to arise from a variety of different orderings, most of which satisfy dominance and weak transitivity, rather than a consistent tendency to violate scalability by a majority of people. Similar results were found for the rating task.

Estimation of Configural Weights

A simple version of configural weight theory (Equations 5a and 5b) was fit to the mean judgments in Tables 1 and 3 as follows. For simplicity, and supported by the findings of Birnbaum et al. (1992) for this range of outcomes, we approximated the utility function for money as linear: $u(x) = x$. The J function in Equation 1 was also assumed to

be linear. Using these assumptions, the weights of the lower valued and higher valued outcomes were estimated to be $b = .553$ and $a = .330$, respectively. Dividing by the sum of the weights, these values yield relative weights of .63 and .37 for the lower and higher valued outcomes. These values equal the relative weights estimated for 50-50 gambles in the neutral point of view by Birnbaum et al. (1992), rounded to the nearest .01. This model correlates only .94 with the mean judgments in Tables 1 and 3, because it cannot describe the violations of scalability shown in Table 3, but it does give a reasonable approximation to the violations of interval independence in Table 1.

Discussion

The violations of interval independence in Tables 1 and 2 provide an important confirmation of configural weighting with strength of preference judgments. They violate expected utility theory in the direction predicted if the lower outcome receives greater weight.

Systematic violations of interval independence are inconsistent with the theory that people edit common features when comparing gambles, as suggested by Kahneman and Tversky (1979). If judges had edited outcomes that are the same in both gambles, then there would have been no systematic effect of the common outcome on their judgments of strength of preference.

The present data with two-outcome gambles indicate that the lower valued outcome receives greater relative weight (.63) than the higher of two positive outcomes (.37), the same as found for judgments of the value of gambles from the neutral's point of view (Birnbaum et al., 1992). This conclusion is not consistent with the particular rank-dependent utility theory of Quiggin (1982), which assumes that the weight of two equally likely alternatives would each be 1/2.

Cumulative prospect theory is a more general rank-dependent model than Quiggin's. As fit by Tversky and Kahneman (1992) to choice-based certainty equivalents of two-outcome gambles, their model implies weights of lower and higher outcomes of .58 and .42, respectively. Cumulative prospect theory also uses a power function for the utility of money, $u(x) = x^{.88}$, which accounts for a portion

of the risk aversion in that theory. To compare results, it is necessary to adjust for the power function. Adjusting for the power function, the effective relative weights of Tversky and Kahneman (1992) would be .63 and .37 for the lower and higher outcomes, respectively, the same as in the present study. [In their model, certainty equivalents of 50–50 gambles of the form $(\$0, x)$ are given by the expression $CE = (.42x^{.88})^{(1/.88)} = x(.42)^{(1/.88)} = .37x$; for example, they reported a median certainty equivalent for the gamble $(\$0, \$100)$ to be \$36.]

In summary, three experiments agree: Birnbaum et al. (1992) obtained judgments of the “fair price” of gambles from the neutral’s point of view and found weights of .63 and .37 for the lower and higher of two positive values; Tversky and Kahneman (1992) found comparable values based on choices between gambles and certain cash; and the present results for judgments of strength of preference between two binary outcome gambles also imply the same weights.

The agreement among the present results for strength of preference, judgments of “fair” prices, and choice-based certainty equivalents should not be taken for granted. In many situations, such as the classic preference reversal (Birnbaum, 1992a; Bostic, Herrnstein, & Luce, 1990; Lichtenstein & Slovic, 1971; Lindman, 1971; Mellers, Chang, Birnbaum, & Ordóñez, 1992; Mellers, Ordóñez, & Birnbaum, 1992; Slovic, Lichtenstein, & Fischhoff, 1988; Stevenson, Busemeyer, & Naylor, 1991; Tversky, Sattath, & Slovic, 1988; von Winterfeldt & Edwards, 1986) and in research on monotonicity violations (Birnbaum, 1992a, 1992b; Birnbaum & Sutton, 1992; Birnbaum & Thompson, 1996; von Winterfeldt, Chung, Luce, & Cho, 1997), choices, preferences, and judgments have not agreed. In certain situations, people may assign higher prices to gamble A than gamble B , but they prefer B to A in a direct comparison. It is not a foregone conclusion that differences in judgment are monotonically related to judgments of difference, nor do judged certainty equivalents of gambles necessarily equal certainty equivalents based on choice.

In the present tasks, judges were not asked to evaluate each gamble separately, as they were in judgment studies such as Birnbaum et al. (1992). Instead, the present results require the judge to rate directly the strength of preference between two gambles or to judge the amount that they would pay to move from one gamble to another. In this case, the evidence of configural weighting obtained in judgments is confirmed, because the judged strength of preference between two outcomes is greater when the common outcome is the highest than when the common outcome is the lowest.

Analogous Findings

The present results are directly analogous to the findings of Birnbaum (1974) who asked people to judge strengths of preference between hypothetical persons, each of whom was described by two adjectives, each provided by a source who knew the person well. Birnbaum found that if one

adjective was unfavorable, the difference due to variation of the other trait was less than if the common adjective was favorable. For example, judges rated the differences in likeableness between the following pairs of persons: (LOYAL & UNDERSTANDING) versus (LOYAL & OBNOXIOUS) and (MALICIOUS & UNDERSTANDING) versus (MALICIOUS & OBNOXIOUS). Birnbaum (1974) found that the first difference in liking was judged to be twice as great as the second difference; this pattern was also replicated with six different sets of adjectives. This finding is comparable to the results in Tables 1 and 2; If the common stimulus component is lower in value (MALICIOUS), the difference due to the other stimulus contrast is less than if the common component is higher in value (LOYAL). In the analogy to Table 1, lower valued outcomes correspond to less likeable traits (Birnbaum, 1982).

Alternative Interpretations of Strength of Preference

It is interesting to consider a ratio interpretation as an alternative to Expression 1:

$$D(A; B) = J \left[\frac{U(A)}{U(B)} \right], \tag{7}$$

where strength of preference is interpreted as a monotonic function of the ratio of utilities, rather than as a difference in utilities. This ratio interpretation, given the data of Tables 1 and 2, also requires rejection of Equation 2. To understand how the ratio interpretation also refutes the additive model, consider the utilities of four gambles such that $U(A) > U(B) > U(D) > 0$ and $U(A) > U(C) > U(D) > 0$. For example, let $A = (\$74, \$100)$, $B = (\$8, \$100)$, $C = (\$6, \$74)$, and $D = (\$6, \$8)$. Suppose the data indicate that $D(A, B) > D(C, D)$, consistent with the pattern of Tables 1 and 2. Equation 7 interprets this result to imply the following:

$$\frac{U(A)}{U(B)} > \frac{U(C)}{U(D)}$$

Subtracting 1 from both sides, we have the following:

$$\frac{U(A)}{U(B)} - \frac{U(B)}{U(B)} > \frac{U(C)}{U(D)} - \frac{U(D)}{U(D)}$$

$$\left[\frac{1}{U(B)} \right] [U(A) - U(B)] > \left[\frac{1}{U(D)} \right] [U(C) - U(D)].$$

Because $U(B) > U(D)$,

$$\left[\frac{1}{U(D)} \right] > \left[\frac{1}{U(B)} \right],$$

therefore:

$$\begin{aligned} \left[\frac{1}{U(D)} \right] [U(A) - U(B)] &> \left[\frac{1}{U(B)} \right] [U(A) - U(B)] \\ &> \left[\frac{1}{U(D)} \right] [U(C) - U(D)]. \end{aligned}$$

Therefore:

$$[U(A) - U(B)] > [U(C) - U(D)].$$

This derivation shows that if $D(A, B) > D(C, D)$, then the ratio interpretation also implies the same ordering on the intervals and therefore provides the same qualitative conclusion on the ordering of utility intervals that refutes the additive model.

There is another argument that could complicate the interpretation of the main result in Tables 1 and 2. It could be argued that a transformation precedes the comparison of the gambles, as in the following equation:

$$D(A; B) = J[T\{U(A)\} - T\{U(B)\}], \quad (8)$$

where T is the strictly monotonic transformation. (Equation 7 is a special case of Equation 8 in which T is the logarithmic function.) In the present experiment, the transformation must be positively accelerated to save the additive model of Equation 2.¹ For Table 1, T would be interpreted as the inverse of the utility function for money. The interpretation would be that prior to comparing two gambles in the present task, one converts their utilities into money. The conversion to money takes the judgment out of the utility scale. The opposite of this argument has been made independently by Eisler (1978) and by Marks (1979), who argued that a negatively accelerated, logarithmic transformation may precede a subtractive operation. (For the situations they discussed, the logarithmic transformation is needed to save a ratio model for psychophysical judgments of "ratios".)

As Birnbaum (1982) noted, these transformation theories lack parsimony, because they require different transformations in each situation without providing new predictions. The ratings in the present experiment also show the same trend in Table 2 as the monetary judgments in Table 1. The positively accelerated transformation, T , that would be required seems less attractive for ratings of preferences (Table 2) than it does for the monetary scale.

Violations of scalability in Table 3 constitute evidence against Equation 1 for gambles selected according to the design used in that table. In defense of Equation 1, however, it should be remembered that Design 2 was planned to create violations of scalability by mixing dominance with preferences for higher expected values versus less risky gambles. The design included four gambles with a dominance relationship and four that traded off expected value against range ($|x - y|$).

Judgments of strength of preference seem to be influenced by certainty of preference as well as magnitude of preference. Similar results have been obtained by Birnbaum and McIntosh (1996) with judgments of strength of preference between gambles with three equally likely outcomes. They concluded that people place a premium on dominance. Indeed, Table 3 shows that the mean judgments differ by more than the expected value difference when one gamble dominates another.

It is also important to note that the major result in Table 1 showed a systematic pattern characteristic of more than 80% of the individuals tested, whereas the violations of

scalability in the means of Table 3, while interpretable, are partially due to averaging different preference orders over different individuals, and they appear not to be as consistent at the individual level as the violations of interval independence. For further discussion of related issues concerning violations of weak and strong transitivity with gambles, see Birnbaum & Thompson (1996), Busemeyer (1985), Stevenson et al. (1991), Tversky (1969), and Mellers and Biagini (1994).

Arguments to the contrary notwithstanding, if we can accept the premise that greater judgments of strength of preference correspond to greater differences in utility for the type of stimuli used in Tables 1 and 2, the present data provide evidence in favor of configural weighting as the explanation for risk aversion. People are willing to pay only about half as much to get (\$6, \$92) instead of (\$6, \$8) than they are to get (\$100, \$92) instead of (\$100, \$8); furthermore, they rate the degree of preference in the former case to be less than in the latter. The simplest interpretation is that the lowest outcome in each gamble receives greater configural weight than the highest outcome in the gamble. Therefore, comparisons between gambles that differ in their lowest outcome produce bigger differences in utility than comparisons between gambles that differ in their higher outcomes.

Converging Evidence of Configural Weighting

These results add another piece to an emerging mosaic of evidence favoring configural weighting over utility as the explanation of risk aversion. Birnbaum and Sutton (1992) found that using configural weight theory leads to scales of utility in better agreement (than nonconfigural theories) with riskless judgments of "ratios" and "differences" in monetary value as well as describing preference reversals due to changes in point of view. Agreement of scales derived from configural weight theory to scales derived from "difference" judgments in other judgment domains has also been found by Birnbaum (1974), Birnbaum and Veit (1974), and Birnbaum and Sotoodeh (1991).

Birnbaum et al. (1992) found that judgments of value violate branch independence and monotonicity in ways that can be explained by configural weight theory (see also Birnbaum, 1992b; Birnbaum & Thompson, 1996; Weber, Anderson, & Birnbaum, 1992). They also found that nonconfigural theories would require different utility functions to explain the finding that judgments from different points of view are not monotonically related (see also Birnbaum & Stegner, 1979).

Branch independence is a weaker form of Savage's "sure thing" principle that asserts that if two gambles have a common branch (if the two gambles have the same outcome produced by the same event of known probability), then the value of the outcome on that branch should not affect the

¹ This type of argument was also made in reference to Birnbaum's (1974) "scale-free" test by Eric Holman (1972, personal communication).

order induced by the other outcomes. For gambles composed of three equally likely outcomes (x, y, z) , branch independence can be written (x, y, z) preferred to (x', y', z) if and only if (x, y, z') preferred to (x', y', z') .

Branch independence is similar to interval independence in that it is implied by a weighted additive or averaging model (although it cannot be tested in full form with binary gambles); however, the test does not require assumptions comparable to Equation 1, but only requires the direction of preference.

Birnbaum and McIntosh (1996) found violations of branch independence in two choice tasks. Most judges preferred $(\$2, \$40, \$44)$ to $(\$2, \$10, \$98)$, but most judges preferred $(\$10, \$98, \$136)$ to $(\$40, \$44, \$136)$. The Birnbaum and McIntosh (1996) test of branch independence does not depend on the assumption of Equation 1; it only requires the direction of preference rather than the strength or magnitude of preference.

Birnbaum and Beeghley (1997) found that judgments of buying and selling prices violate branch independence and do so in ways that can be predicted from a changing pattern of configural weights, ruling out nonconfigural weighted utility theories including even those in which the utility changes in different points of view.

These studies show consistently that in the buyer's perspective and in choice situations people assign greater weight to lower valued stimuli. In the seller's task, greater weight is assigned to the middle stimulus, producing new violations of branch independence. All three tasks (buying prices, selling prices, and choices) can be predicted using the same utility function, even though all three invoked different preference orders for the same gambles. The different orders were explained by different configural weights in the three tasks. Results also showed that buying, selling, and choice tasks had weights consistent with the property that the ratio of the weight of the medium outcome to the weight of the highest outcome exceeds the ratio of the weight of the lowest outcome to the medium. As shown by Birnbaum and McIntosh (1996), this pattern of ratios of weights implies a particular systematic violation of branch independence.

Cumulative prospect theory, however, as fit by Tversky and Kahneman (1992), implies that for gambles composed of three equally likely outcomes, the weight of the middle outcome should be less than the weight of either the lowest or the highest outcome. Because of that implication, the weighting function of Tversky and Kahneman (1992) predicts the opposite pattern of violation of branch independence from what is observed in both judgment (Birnbaum & Beeghley, 1997) and choice (Birnbaum & McIntosh, 1996) experiments. Because configural weight theory can explain both the results of Tversky and Kahneman (1992) for two outcome gambles and newer results with three-outcome gambles, it appears preferable to the model of Tversky and Kahneman (1992).

Putting the present piece of the mosaic into the picture produced by other studies, the present results show evidence of violation of interval independence in favor of the theory that lower outcomes get more weight than higher outcomes.

In this finding, results for strength of preferences agree with those of judgments of price, with choice-based certainty equivalents for two-outcome gambles, and with evidence from choices between three outcome gambles. These results show that people do not edit common outcomes when comparing gambles but instead weight the contrasting outcomes less when the common outcome is low than they do when the common outcome is high in value.

References

- Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology*, *102*, 543-561.
- Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401-485). Hillsdale, NJ: Erlbaum.
- Birnbaum, M. H. (1992a). Issues in utility measurement. *Organizational Behavior and Human Decision Processes*, *52*, 319-330.
- Birnbaum, M. H. (1992b). Violations of monotonicity and contextual effects in choice-based certainty equivalents. *Psychological Science*, *3*, 310-314.
- Birnbaum, M. H., & Beeghley, D. (1997). Violations of branch independence in judgments of the value of gambles. *Psychological Science*, *8*, 87-94.
- Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 331-346.
- Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes*, *67*, 91-110.
- Birnbaum, M. H., & Sotoodeh, Y. (1991). Measurement of stress: Scaling the magnitudes of life changes. *Psychological Science*, *2*, 236-243.
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*, 48-74.
- Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes*, *52*, 183-215.
- Birnbaum, M. H., & Thompson, L. A. (1996). Violations of monotonicity in choices between gambles and certain cash. *American Journal of Psychology*, *109*, 501-523.
- Birnbaum, M. H., & Veit, C. T. (1974). Scale-free tests of an averaging model for the size-weight illusion. *Perception & Psychophysics*, *16*, 276-282.
- Bostic, R., Herrnstein, R. J., & Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifference. *Journal of Economic Behavior and Organization*, *13*, 193-212.
- Busemeyer, J. R. (1985). Decision making under uncertainty: Simple scalability, fixed sample, and sequential sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 583-564.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380-417.
- Eisler, H. (1978). On the ability to estimate differences: A note on Birnbaum's subtractive model. *Perception & Psychophysics*, *24*, 185-189.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.
- Lichtenstein, S., & Slovic, P. (1971). Reversals of preference

- between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89, 46–55.
- Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology*, 89, 390–397.
- Lopes, L. (1990). Re-modeling risk aversion: A comparison of Bernoullian and rank dependent value approaches. In G. M. v. Furstenberg (Ed.), *Acting under uncertainty* (pp. 267–299). Boston: Kluwer.
- Luce, R. D. (1992). Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty*, 5, 5–27.
- Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty*, 4, 29–59.
- Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology*, 29, 1–72.
- Marks, L. E. (1979). A theory of loudness and loudness judgments. *Psychological Review*, 86, 256–285.
- Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review*, 101, 505–518.
- Mellers, B. A., Chang, S., Birnbaum, M. H., & Ordóñez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 347–361.
- Mellers, B. A., Ordóñez, L., & Birnbaum, M. H. (1992). A change of process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes*, 52, 331–369.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3, 324–345.
- Rose, B. J., & Birnbaum, M. H. (1975). Judgments of differences and ratios of numerals. *Perception & Psychophysics*, 18, 194–200.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Slovic, P., Lichtenstein, S., & Fischhoff, B. (1988). Decision making. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (Vol. 2, pp. 673–738). New York: Wiley.
- Stevenson, M. K. (1986). A discounting model for decisions with delayed positive or negative outcomes. *Journal of Experimental Psychology: General*, 115, 131–154.
- Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1991). Judgment and decision-making theory. In M. Dunnette & L. M. Hough (Eds.), *New handbook of industrial-organizational psychology* (Vol. 1, pp. 283–374). Palo Alto, CA: Consulting Psychologists Press.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371–384.
- von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.
- von Winterfeldt, D., Chung, N., Luce, R. D., & Cho, Y. (1997). Tests of consequence monotonicity in decision making under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 406–426.
- von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, England: Cambridge University Press.
- Wakker, P. (1993). Additive representations on rank-ordered sets, II: The topological approach. *Journal of Mathematical Economics*, 22, 1–26.
- Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. *Theory and Decision*, 36, 1–44.
- Wakker, P., Erev, I., & Weber, E. U. (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty*, 9, 195–230.
- Weber, E. U. (1994). From subjective probabilities to decision weights: The effects of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, 114, 228–242.
- Weber, E. U., Anderson, C. J., & Birnbaum, M. H. (1992). A theory of perceived risk and attractiveness. *Organizational Behavior and Human Decision Processes*, 52, 492–523.

Received September 18, 1995

Revision received March 21, 1996

Accepted March 21, 1996 ■