Choice, Decision, and Measurement
A. A. J. Marley (Ed.),

VIOLATIONS OF MONOTONICITY IN JUDGMENT AND DECISION MAKING

MICHAEL H. BIRNBAUM

California State University, Fullerton
and Institute for Mathematical Behavioral Sciences, Irvine

ABSTRACT. Monotonicity is a fundamental assumption of axiomatic decision theories. According to this principle, if two alternatives are otherwise identical, except one has an outcome for at least one nonempty state of nature that is preferred to the corresponding outcome for the other alternative, then the alternative with the better outcome should be preferred. Applied to judgments of the value of gambles, the principle states that the judged value of a gamble should increase monotonically as a function of each outcome, holding everything else constant. As appealing as this axiom is for normative theory, it has been systematically violated in experiments in which subjects judge cash values of gambles. The violation has not been observed in transparent, direct comparisons, but it has been replicated when the gambles are compared to a fixed set of cash values. The violations can be explained by the assumption that decision weights in judgment differ depending on the rank and also on the augmented sign (which is negative, zero, or positive). Violations of branch independence can also be explained by rank-dependent configural weighting. The pattern observed rules out the theory that subjects cancel common outcomes in comparison. The pattern is also opposite that predicted by the inverse-S weighting function used in cumulative prospect theory. Testable properties are suggested to distinguish different models of configural weighting.

1. INTRODUCTION

The principle of consequence monotonicity can be stated briefly as follows: If two alternatives are otherwise identical but one alternative has a consequence for one nonempty state of nature that is preferred to the corresponding consequence for that state of nature given the other alternative, then the alternative with the better consequence should be preferred.

For gambles defined as probability distributions whose consequences are monetary outcomes, outcome monotonicity can be defined as follows:

Key words and phrases. Decision making and judgment, axiomatic theories of decision-making, monotonicity, Savage's axiom, subjective expected utility theory, cumulative prospect theory, configural weighting, rank-dependent utility theory, dominance, choice, preference, utility.

Acknowledgments. Research summarized in this paper was supported by a National Science Foundation Grant, SES 8921880, to California State University, Fullerton Foundation. Preparation of this chapter was facilitated by SBR-9410572. I thank R. Duncan Luce, Barbara Mellers, John Miyamoto, Peter Wakker, Elke Weber, and an anonymous reviewer for helpful comments on an earlier draft.

Address for correspondence. Prof. Michael H. Birnbaum, Department of Psychology, California State University, P.O. Box 6846, Fullerton, CA 92834-6846. E-mail: mbirnbaum@fullerton.edu

Suppose gambles A and A' differ in their outcomes on one branch as follows:

$$A = (x, p(x); a_2, p(a_2); \ldots; a_i, p(a_i); \ldots; a_m, p(a_m))$$
$$A' = (y, p(x); a_2, p(a_2); \ldots; a_i, p(a_i); \ldots; a_m, p(a_m))$$

where $p(x)$ is the probability to receive outcome x (or y), given choice A (or A'), respectively; and the sum of the probabilities is 1 within each gamble. Monotonicity requires that gamble A is preferred to A' if and only if gamble B is preferred to B' where:

$$B = (x, p'(x); b_2, p(b_2); \ldots; b_j, p(b_j); \ldots; b_n, p(b_n))$$
$$B' = (y, p'(x); b_2, p(b_2); \ldots; b_j, p(b_j); \ldots; b_n, p(b_n))$$

for all gambles so defined.

The term *stochastic dominance* refers to the relation between nonidentical gambles, A and B, such that gamble A stochastically dominates gamble B if and only if the probability of receiving x or less given gamble A is less than or equal the probability of receiving x or less given gamble B, for all x. Tversky and Kahneman (1986) reported a violation of stochastic dominance, when the relation was not transparent.

Stochastic dominance combines monotonicity with respect to outcomes and monotonicity with respect to probabilities. A violation of outcome monotonicity also violates stochastic dominance, but a violation of stochastic dominance is not necessarily a violation of monotonicity, unless other assumptions are made (Luce, 1986a, 1988). Luce and von Winterfeldt (1994) noted that it is therefore useful to decompose the concept of dominance into consequence (outcome) monotonicity and event monotonicity. In Luce's (1988) approach, the Ellsberg paradox can be interpreted as a violation of event monotonicity, but is not a test of outcome monotonicity.

This chapter deals with outcome monotonicity, concerning which Luce (1992c) remarked, "Because monotonicity is a keystone to all existing theories of choices among uncertain alternatives, it is essential that we decide whether or not it is generally applicable. If not, it's back to the drawing boards." (*p. 23*)

The next section reviews research showing that certain types of judgments systematically violate outcome monotonicity. A configural weight model is presented in the third section to describe the violations. The fourth section presents experimental replications and extensions of the research paradigm. The fifth section takes up choice-based certainty equivalents, which give mixed results, depending on the method used. The sixth section reviews experiments that have estimated the configural weighting function for positive, negative, and zero outcomes. The seventh section takes up a related phenomenon, violations of branch independence, which can be used to test among different configural weighting models. The eighth section summarizes the current status of evidence and describes testable properties that can be used to compare different classes of configural weighting models, and the ninth section gives a summary of conclusions.

2. Violations of Monotonicity in Judgment

Although outcome monotonicity seems a very reasonable axiom for the rational decision maker, recent experiments have found situations in which mean judgments violate the principle systematically. Birnbaum and Gregory Coffey designed and conducted two experiments in 1986, following the approach of Birnbaum and Stegner (1979). Their first experiment showed violations of monotonicity that appeared to indicate that the outcome of zero receives less weight than nonzero outcomes, similar to results previously reported by Anderson and Birnbaum (1976). Birnbaum and Coffey designed a stronger test for violations in their second experiment. Their results were reported by Birnbaum (1987a, 1987b, 1987c).

At this time, collaborative projects were under way with Elke Weber, Barbara Mellers, Carolyn Anderson, and Lisa Ordóñez to investigate whether principles of judgment, inferred from judgments in other domains, also applied to judgments of gambles. Weber et al. (1992) applied the approach of Birnbaum and Stegner (1979) to model the relationships between ratings of the risk and attractiveness of gambles. One line of research was devoted to testing the models of preference reversals of Goldstein and Einhorn (1987) and Tversky, Sattath, and Slovic (1988), using the criterion of scale convergence (Birnbaum, 1974, 1982). Although the expression theory of Goldstein and Einhorn (1987) can accommodate violations of monotonicity, neither it nor the contingent weighting theory of Tversky et al. (1988) correctly accounts for changes in rank order between ratings and prices using the same scale of utility (Mellers, Ordóñez, & Birnbaum, 1992a).

The experiments of Birnbaum and Coffey had been designed to test several predictions of configural weighting models, including a specific pattern of changes in rank order of judgments that should be produced by point of view if point of view affects configural weights, as postulated in Birnbaum and Stegner (1979). The violations of monotonicity, confirmed in their second experiment, excited interest and became the focus of new research.

Sara Sutton, Barbara Mellers, Patricia Berretty, and Robin Weiss soon joined in the study of these phenomena, fitting the data to models and exploring the effects of different subjects, different values, different stimulus formats, and other variations. Results of these investigations were published in several papers (Birnbaum, 1992b; Birnbaum et al., 1992; Birnbaum & Sutton, 1992; Mellers et al., 1992b; Weber et al., 1992). That research led to further investigations (Birnbaum & Thompson, 1996; Mellers et al., 1995; von Winterfeldt et al., 1997). This chapter will review this body of empirical research.

Let (x, p_x, y) represent the binary gamble to receive x with probability p_x and otherwise receive y ($p_y = 1 - p_x > 0$). Monotonicity requires that (x_1, p_1, y) is preferred to (x_2, p_1, y) if and only if (x_1, p_2, y) is preferred to (x_2, p_2, y); in other words, if and only if x_1 is preferred to x_2. Birnbaum et al. (1992) found that ($0, .05, $96) receives a higher mean judgment than ($24, .05, $96), although ($24, .5, $96) receives a higher mean judgment than ($0, .5, $96); indeed, we assume that $24 is better than $0.

Figure 1 illustrates the pattern of results observed by Birnbaum et al. (1992). Mean judgments from the buyer's point of view (the "most a buyer should pay" for each gamble), the neutral's point of view ("fair" price), and the seller's point of view (the "least a seller should accept to sell the gamble") are plotted in separate panels as a function of the probability to win $96, with open circles for gambles in which the lowest outcome was $0, and filled circles for gambles in which the worst outcome was $24. According to outcome monotonicity, filled symbols ($24) should exceed the open symbols ($0), for all values of p. Instead, the data values cross for three levels of $1-p \geq .8$ in each point of view. The percentage of subjects who violated monotonicity when $1-p=.95$ was 53%, 60%, and 36% in the buyer's, neutral's and seller's points of view, compared with 34%, 25%, and 36% who conformed to it, respectively, and the rest were ties. Similar results were observed when $72 replaced $96.

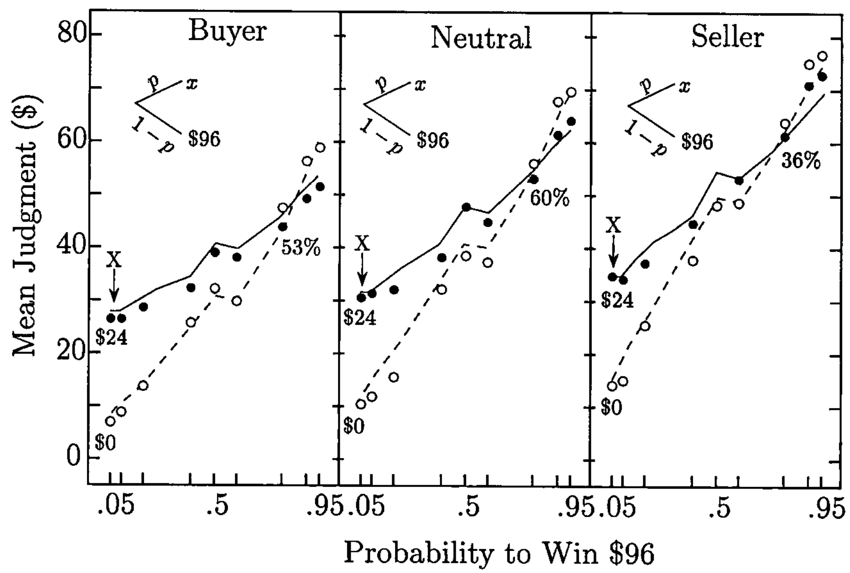

FIGURE 1. Mean judgments of binary gambles (x, p, y) as a function of the probability to win $96 $(1-p)$, with unfilled circles showing results when $x=$ $0, and filled circles showing results when $x=$ $24. Separate panels show judgments in the buyer's, neutral's, and seller's points of view. Crossing of the curves indicates violation of outcome monotonicity. Lines show predictions of configural weight model. Data from Birnbaum et al. (1992).

3. Configural Weight Model of Violations and Point of View

The term *configural* is used to indicate that the parameter representing a stimulus component depends on the relationships between that component and others that comprise the stimulus array presented on each trial (Birnbaum, 1974). Subjective expected utility (SEU) theory (Savage, 1954), for example, is not configural

because the weight of each outcome's utility is independent of the value of the outcome and its relationships to other outcomes in the same gamble, and the utility of each outcome is independent of the other outcomes. Similarly, Edwards (1954) version of SEU using a weighting function for probabilities is also not configural, for the same reason. "Configural weighting" models allow the weights of the outcomes to depend on the configuration of outcomes and probabilities that comprise the gambles, but assume that the utility function is independent of context and configuration.

Birnbaum's (1974, *p. 559*) configural weight model allows the weight of a stimulus component to depend on its rank among the other components that comprise the stimulus array. Applied to gambles, the weight of the same outcome with the same probability can be different in different gambles depending on the other outcomes in those gambles (Birnbaum, 1982, 1992a, 1992b; Birnbaum et al., 1992; Birnbaum & Sotoodeh, 1991; Birnbaum & Stegner, 1979; Weber et al., 1992; Weber, 1994).

Configural weighting models, such as Birnbaum's (1974) range model, are closely related to rank-and sign-dependent utility models (Chew & Wakker, 1996; Lopes, 1990; Luce, 1992b, 1992c, 1995b; Luce & Fishburn, 1991, 1995; Luce & Narens, 1985; Quiggin, 1982; Tversky & Kahneman, 1992; Yaari, 1987), which were developed independently [see review by Wakker (1993)].

Rank-dependent utility theory and rank-and sign-dependent utility theories are configural weight models that allow violations of outcome independence (see Section 7), but assume monotonicity. The model of cumulative prospect theory is a special case of rank- and sign-dependent theory with a restricted weighting function; this model implies stochastic dominance. To account for violations of monotonicity, the numerical representation of rank- and sign-dependent utility theory (Luce & Fishburn, 1991), for example, would have to be modified to allow different weights for different outcomes. Luce (1992b) noted that the violations observed thus far have been restricted to gambles including the outcome zero, and suggested how a rank- and sign-dependent representation of certainty equivalents could be modified to accommodate the violations. In the eighth section of this chapter, empirical properties are described that can test among various configural weight models. In this section, we present the model of Birnbaum et al. (1992) to account for violations of monotonicity and which also describes changes in rank order that depend on the judge's point of view.

Birnbaum et al. (1992) represented judgments of binary gambles, (x, p_x, y), by the following configural weight model:

$$U_V(x, p_x, y) = \frac{Au(x) + Bu(y)}{A+B} \tag{1a}$$

where $U_V(x, p_x, y)$ is the utility of the gamble in point of view V; $u(x)$ and $u(y)$ are the utilities (subjective values) of the lower- and higher-valued outcomes $(x < y)$; and A and B are their absolute configural weights, which depend on the judge's point of view, on probability, and value as in the following equations:

$$A = a_V S_x(p_x) \tag{1b}$$

$$B = (1-a_V)[1 - S_x(1-p_y)] \tag{1c}$$

where a_V is the configural weighting parameter for point of view V; p_x and p_y are the probabilities to receive x or y, respectively; and $S_x(p_x)$ is a function of the probability of the lower-valued outcome, x, that depends on its value.

Birnbaum et al. (1992) posited two different S_x functions for the cases in which $x > 0$ and for $x = 0$. For $.04 < p < .96$, $S_x(p)$ can be approximated by $S_x(p) = .59p + .29$, for $x > 0$; however, for $x = 0$, $S_0(p)$ is approximated by $S_0(p) = .74p + .14$. Note that $S_0(p)$ is less than $S_x(p)$, especially for small values of p. In this model, monotonicity violations occur because the lowest outcome of zero has less weight than a lowest outcome that is a small positive amount (for the same low probability).

For three-outcome gambles $(x, p_x; y, p_y; z, p_z)$, where $0 \leq x < y < z$, and $p_z = 1 - p_x - p_y$, Birnbaum et al. (1992) used the following expression:

$$U_V(x, p_x; y, p_y; z, p_z) = \frac{Au(x) + Bu(y) + Cu(z)}{A + B + C} \tag{1d}$$

where

$$C = (1 - a_V)[1 - S_x(1 - p_z)]. \tag{1e}$$

A and B are as defined in Expressions 1b and 1c.

In this model, $u(x)$ and $S_x(p_x)$ are assumed to be independent of point of view, context, and configuration. Birnbaum et al. (1992) assumed that the weights of the middle and highest stimuli are equal when they are of equal probability (i.e., $C = B$ when $p_y = p_z$), an assumption that will be reconsidered in Section 7.

This model can be derived from the assumption that the subject is minimizing an asymmetric loss function, assuming that the stimuli are spaced so that the response is between the lowest and middle stimuli (Birnbaum et al., 1992; Birnbaum & McIntosh, 1996). When the response is between the middle and highest stimuli, however, the weight of the middle stimulus would equal that of the lowest outcome, a switch of configural weights that allows violations of comonotonic independence (Birnbaum & McIntosh, 1996). Thus, the loss function approach implies that configural weights will depend not only on rank but also on the spacing of the stimuli. Experiments to test these interpretations are proposed in Section 8.

The loss function concept also provides a rationale to explain why configural weights would depend on the judge's point of view (Birnbaum & Stegner, 1979). Judge's "point of view" refers to instructions that may affect the relative costs of judgment errors in different directions. Examples of viewpoint manipulations are instructions to the judge to identify with the buyer or seller in a transaction, to judge the morality of others or to consider being judged, or to identify with the prosecution or defense in a trial. If point of view affects the relative costs of over- or under-estimating a value, and if judges choose responses to minimize costs, then configural weights should depend on viewpoint. If weights change in different viewpoints, the models predict special patterns of reversals of preference due to changes in point of view (Birnbaum, 1982; Birnbaum & Beeghley, 1997).

Previous research that fit configural weight models to judgments concluded that weights also differ for neutral, or zero-valued outcomes (Anderson & Birnbaum, 1976); such an assumption allows configural weighting to explain violations of monotonicity. Configural weight models assume scale convergence, the principle that the utility (or value) function is independent of point of view and configuration. The assumption of scale convergence was used to test rank dependent models against the nonconfigural models (Birnbaum et al., 1992; Birnbaum & Sutton, 1992).

The configural weight parameters, a_V, predict how the rank order of gambles change in different points of view. For the seller's point of view, a_V was set to .5, and the values estimated for the neutral's and buyer's points of view were approximately .6 and .7, respectively. Configural weight theory (Expressions 1a-1e) led to an estimated $u(x)$ function that was invariant with respect to point of view (Birnbaum et al., 1992); nonconfigural theories require different $u(x)$ functions in different viewpoints. This model led to an estimated $u(x)$ function that was also compatible with estimates of $u(x)$ based on judgments of "ratios" and "differences" of riskless utility (Birnbaum & Sutton, 1992).

4. Replications and Extensions

Monotonicity Satisfied in Direct Comparisons. Birnbaum and Sutton (1992), as part of their study of scale convergence, included a partial replication of the tests of monotonicity from Birnbaum et al. (1992). They also asked subjects to choose between pairs of gambles, including pairs involving tests of monotonicity. Although mean and median judgments violated monotonicity, replicating the findings of Birnbaum et al. (1992), few subjects violated monotonicity when asked to make direct comparisons.

Figure 2 shows mean judgments from Birnbaum and Sutton (1992) in the seller's point of view. [Medians for both buyer's and seller's viewpoints are similar, as shown in Birnbaum & Sutton (1992, Figure 9)]. Mean judgments for ($0, p, $96) and for ($0, p, $72) are shown as open squares and circles, respectively, connected by dashed lines; mean judgments of ($24, p, $96) and ($24, p, $72) are shown as filled squares and circles, respectively, connected by solid lines. The crossing of open and filled symbols represent violations of monotonicity. Figure 2 also shows that judgments of ($24, p, $96) are not simply $24 plus the judgments of ($0, p, $72).

Equations 1a-1e fit the data of Birnbaum et al. (1992), and predicted the patterns of monotonicity violations obtained by Birnbaum and Sutton (1992) and Birnbaum (1992b). The gambles in these studies were presented as in the following example:

.2	.8
$24	$96

which represents ($24, .2, $96). One possibility was that subjects were violating monotonicity because of some numerical algorithm induced by the particular stimulus display.

Violations with Pie charts, Negative Outcomes, and Cash Incentives. Mellers et al. (1992b) replicated and extended the investigation, using a graphical display of probability, different numerical values, and different subjects. Whereas Birnbaum et al. (1992) and Birnbaum and Sutton (1992) had used numerical probabilities, Mellers et al. (1992b) represented probability by means of pie charts, to see

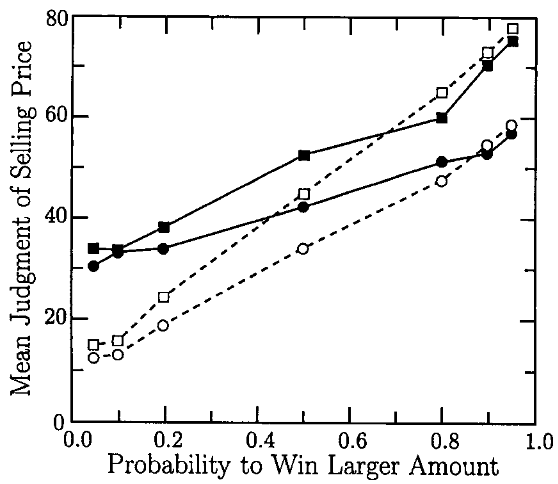

FIGURE 2. Mean judgments of selling prices plotted against the probability to win either \$96 (squares) or \$72 (circles) with unfilled and filled symbols showing results when the lower outcome is \$0 or \$24, respectively. Data from Birnbaum and Sutton (1992).

if the violations would persist when probability was presented graphically. Figure 3 shows an example display.

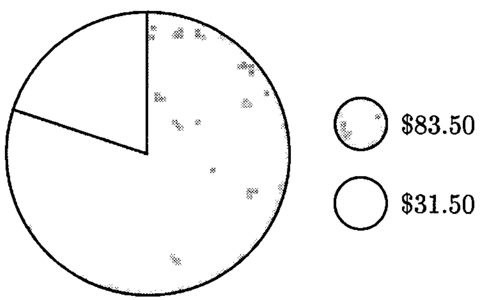

FIGURE 3. An example stimulus display, using a pie chart to represent probability. This stimulus represents (\$31.50, .2, \$83.50)

With these stimuli, similar violations were observed. Mellers et al. (1992b) studied judgments of gambles of the form (x, p, y) and $(0, p, y)$ as a function of x and p. Judgments were made from an ownership point of view, in which the subject judged either the lowest selling price (to give up playing favorable gambles) or the most they would pay (to avoid playing unfavorable gambles, like buying insurance).

Mean judgments are shown in Figure 4, with negative numbers representing offers to pay to avoid the gamble. With y set to \$83.50, judgments are plotted as a function of x with a separate curve for each level of the probability to win

\$83.50 $(1-p)$. A violation of monotonicity is observed for all seven levels of p, comparing mean judgments of $(\$0, p, \$83.50)$, shown as unfilled circles, compared with $(\$5.40, p, \$83.50)$, shown as filled circles connected by dashed lines. (Note that all seven dashed curves have negative slope.) There were also significantly more individual violations of monotonicity between $(\$0, .05, \$83.50)$ and $(\$31.50, .05, \$83.50)$ than between $(\$5.40, .05, \$83.50)$ and $(\$31.50, .05, \$83.50)$, even though the former comparison has a greater difference in expected value.

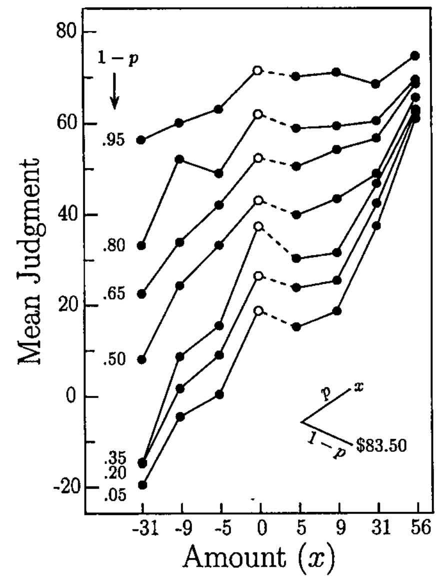

FIGURE 4. Mean judgments of binary gambles of the form $(x, p, \$83.50)$, plotted as a function of x with a separate curve for each level of $1-p$, the probability to win \$83.50. The nonmonotonic "kink" at $x = \$0$, shown as unfilled circles and dashed lines, depicts a violation of monotonicity for each probability used. Data from Mellers, et al. (1992b)

In another experiment, Mellers et al. (1992b) used monetary incentives, instructing subjects that they would play for real cash payoffs one of two selected gambles, the one to which they assigned the higher value. Violations of monotonicity in judgment persisted even when real money was used as an incentive. They also found that violations were observed when both x and y are negative; for example, subjects offered to pay more on the average to *avoid* the gamble $(\$0, .05, -\$85.50)$ than the gamble $(-\$31.50, .05, -\$85.50)$. Similar violations were obtained when the absolute magnitudes of the stakes were changed by multiplying both outcomes by the same constant (see Mellers, et al., 1992b, Figures 6-8 and Tables 1 and 2). Violations were rare, however, when x and y were of opposite sign.

Violations in Three Outcome Gambles. One interpretation of the configural weighting explanation of monotonicity violations was that subjects adopt a simplifying strategy with two-outcome gambles, so zero outcomes would receive a reduced weight only in simple, two-outcome gambles. For these two outcome gambles only (this idea assumes), people multiply probability and value, ignoring the outcome of zero. When there are two nonzero outcomes or three outcomes, they average the outcomes using weights that are more "regressed" than probabilities. This interpretation implies that violations of monotonicity should not occur with three-outcome gambles having two nonzero outcomes. Nevertheless, the same equations and approximated parameters from Birnbaum et al. (1992) successfully predicted violations of monotonicity in a new set of three-outcome gambles (Mellers et al., 1995), using the assumption of Birnbaum et al. (1992) that the lowest outcome receives the same absolute weight in both two- and three-outcome gambles, and the other two outcomes each receive the weight that a higher outcome receives in a two-outcome gamble (Expressions 1a-1e).

In later work with three outcome gambles, this simplifying assumption was revised (as will be discussed in Section 7); nevertheless, the simple assumption and extrapolation of parameters from Birnbaum et al. (1992) to three outcomes did a fair job predicting violations of monotonicity with the new gambles.

Although violations of monotonicity have been found consistently in judgment studies in which the key gambles are judged separately, conditions that facilitate comparisons among the gambles appear to reduce violations. Mellers et al. (1992b) found that when the two gambles involving a dominance relation are presented for judgment in a short list of gambles, the frequency of violations is reduced. Because direct choices yield a different ordering from that obtained from judgment, Birnbaum and Sutton (1992) identified their finding as a new type of preference reversal between judgment and choice.

5. Monotonicity and Choice-Based Certainty Equivalents

The certainty equivalent is the amount of cash that is psychologically indifferent to a gamble. Some preference reversals can be reduced when choice rather than judgment is used to find certainty equivalents (Bostic et al., 1990), so it is reasonable to ask if the choice task itself, rather than the transparency of the choices presented, induces conformity to monotonicity.

Choices between Gambles and Fixed Set of Cash Values. Birnbaum (1992b) offered subjects choices between gambles and a list of cash values that was the same for all gambles. By examining how each gamble stacked up against a fixed set of cash amounts, this procedure separates choice from transparent comparison. Birnbaum (1992b) found that violations of monotonicity persisted even when gambles are ordered by choice-based certainty equivalents (based on comparisons between gambles and a list of sure amounts of money).

The following model is useful for discussing choices between gambles and cash:

$$P(c, \mathrm{G}) = F[u(c) - U(\mathrm{G})] \qquad (2)$$

where $P(c, \mathrm{G})$ is the probability of choosing the sure cash, c, over the gamble $\mathrm{G} = (x, p, y)$; U is a function that assigns an overall utility to each gamble; u is

the utility function for money; F is a monotonic function that maps a given utility difference into a choice probability.

The certainty equivalent, c^*, of gamble G is defined as the value of cash that would be indifferent to the gamble in the sense that it would be preferred half the time; i.e., the value of c^* for which $P(c^*, \mathrm{G}) = 1/2$. Birnbaum (1992b) found that values of c^* violate monotonicity, when certainty equivalents are determined by a choice procedure in which each gamble is compared to a fixed set of comparison cash amounts.

Contextual Effects in Choice. Birnbaum (1992b) also found that the value of c^* depends on the particular set of comparisons used; higher values of c^* are observed when the average value of the cash amounts offered for comparison are higher than when the cash amounts are lower on the average. An example of contextual effects found by Birnbaum (1992b) is illustrated in Figure 5. Note that the inferred certainty equivalent for this gamble, ($0, .95, $48), is larger when the context of comparison cash values has a median of $77 (filled circles) than in the context of comparisons with a median of $14 (open circles). The fact that choice indifference points depend on the context of comparisons makes the interpretation of choice-based certainty equivalents more complicated.

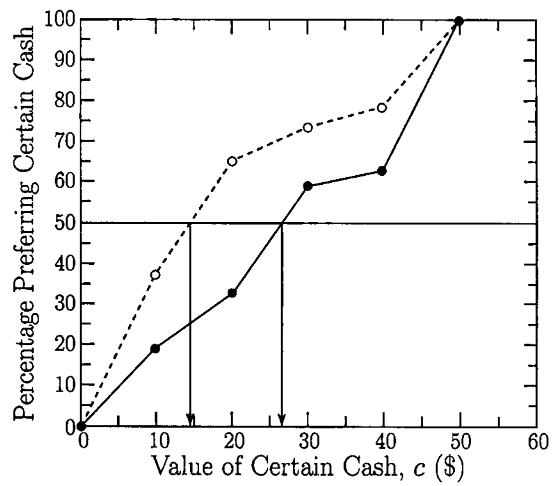

FIGURE 5. Percentage choosing cash over the gamble ($0, .95, $48) as a function of the value of cash, with a separate curve for each context. Each context used a different set of cash values; filled circles show results when median of cash values was $77; open circles show results when median of comparisons was $14. Note that the certainty equivalents (abscissa projections of 50%) are larger when the context of cash values has a higher median. Redrawn from Birnbaum (1992b).

PEST-based Certainty Equivalents. von Winterfeldt et al. (1997) found different rates of violations of monotonicity when certainty equivalents were obtained from different procedures. Using PEST, a staircase method in which different gambles receives different cash comparisons, depending on each subject's choices, they concluded monotonicity violations are less frequent than they are in judgment. Procedures such as PEST confound the distribution of cash values offered with the gambles to be assessed. Because the same gamble can receive different indifference points when different contexts of comparisons are used, when different gambles are presented with different sets of comparisons, it is difficult to know what would have happened if the gambles had been compared to the same standards.

One (overly simple) model of contextual effects is to assume that on some portion of the trials, the subject chooses randomly. If so, then the observed choice indifference point will be a compromise between the "true" choice indifference point and the median of the comparison cash values offered. Thus, the finding by von Winterfeldt et al. (1997) that certainty equivalents in the PEST method obey monotonicity may be due either to the subjective values of the gambles obeying monotonicity, or the fact that on the average higher cash values are presented for comparison to the dominant gambles in this method.

Unfortunately, the PEST algorithm (and the algorithm used by Tversky and Kahneman, 1992, as well) allows a gamble of higher expected value to receive comparisons of higher average value than a gamble of lower expected value. Such a procedure may thus find greater satisfaction of monotonicity because it capitalizes on contextual effects and the monotonicity of expected values rather than because the procedure itself reveals a "truer" measure of certainty equivalents. Although an attempt was made by von Winterfeldt et al. (1997) to statistically correct for differences in context, statistical partialling does not properly correct for confounded variables (Birnbaum & Mellers, 1989).

It would be useful to explore a variation of the PEST procedure using the same values of sure cash for both gambles being compared. One approach would be to study directly contextual effects by systematic variation of the algorithm. Another approach would be to "yoke" two gambles, such as ($0, .05, $96) and ($24, .05, $96), so that the same cash comparisons were presented on different trials against these two gambles. Such an experiment could provide the same context for both gambles being compared.

Scalability and Monotonicity in Choices between Cash and Gambles. Birnbaum and Thompson (1996) considered the following set of relations. For each value of c, operationally define the relation, $\succ_c$, as follows:

$$\mathrm{A} \succ_c \mathrm{B} \text{ if and only if } P(c, \mathrm{A}) < P(c, \mathrm{B}) \tag{3}$$

where $P(c, \mathrm{A})$ represents the proportion of subjects preferring cash amount c over gamble A in a context in which the distribution of cash amounts is fixed for all gambles.

If Equation 2 held with a single function F, then all of the relations in Expression 3 should agree (i.e., the comparison between two gambles would be independent of the cash value c). The agreement of the relations in Expression 3 is termed *scalability*. If F were subscripted for each gamble, then the inferred ordering of gambles in this set of relations can depend on the value of c, violating scalability. Busemeyer (1985) found violations of scalability that suggest that F in Equation 2 depends on the variance of the outcomes within the gamble. Birnbaum and Thompson (1996) found that observed choice proportions violate both monotonicity and scalability.

Figure 6 illustrates these violations by plotting the proportion of choices favoring the cash over ($0, .2, $96) and ($48, .2, $96), shown as open and filled circles, against the value of cash, c. Crossing of curves in Figure 6 represent violations of scalability. Monotonicity is violated in Figure 6 when the open circles are below and to the right of the filled circles. For values of c less than $48 (the lowest positive outcome), monotonicity of $\succ_c$ is satisfied, but when $c >$ $55, it is systematically violated. For these gambles, certainty equivalents (abscissa projections of c^* corresponding to ordinate = 50%) and the $\succ_c$ relationship (for $c >$ $55) violate monotonicity.

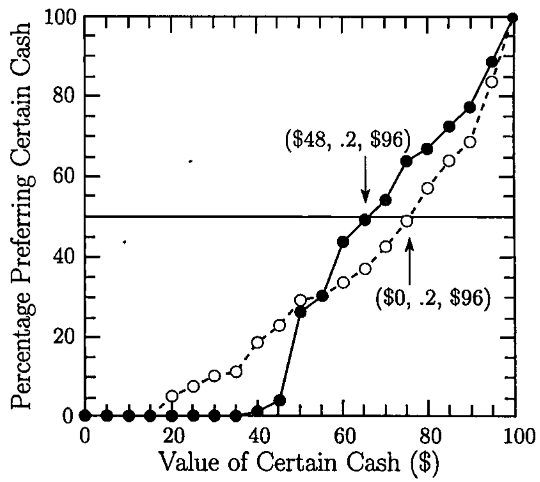

FIGURE 6. Percentage of choices preferring cash to gambles, plotted as a function of the amount of cash, using solid circles for ($48, .2, $96), and open circles for ($0, .2, $96). Crossing of the curves represents violation of scalability. When solid circles are above open circles, there is a violation of monotonicity for that value of c. In this case, monotonicity is satisfied for $c <$ $55, but not for $c >$ $55. Certainty equivalents (abscissa projections of 50%) also violate monotonicity. Similar results were obtained for other gambles. Data from Birnbaum & Thompson (1996).

6. WEIGHTING FUNCTIONS

The violations of outcome monotonicity can be predicted by different weighting functions for gambles with or without the outcome zero. This section explores determinants of the weighting functions as a function of the number, rank, and sign of the outcomes.

Weights of Equally Likely Outcomes as a Function of Number. Birnbaum and McCormick (1991) used yet another procedure for investigating violations of monotonicity in judgment. Their experiment was also designed to estimate weighting functions for positive, negative, and zero outcomes presented with different frequencies. Gambles were presented in the form of a list of equally likely outcomes that would be placed into an urn, from which one would be drawn at random to determine the outcome. For example,

$$(\$24, \$96, \$96, \$96, \$96)$$

represents an urn with five equally-likely tickets, from which one will be selected at random to determine the prize. This gamble offers a probability of .2 to win \$24 and a .8 probability to win \$96. However, in this procedure, the probabilities are not stated, but left to the subject to infer from the lists of values.

Gambles were judged from the viewpoint of "receipt indifference". Subjects were instructed to judge the amount of money that was equal to each gamble in the sense that they would be indifferent between receiving (or paying) that amount or receiving (or paying out) the outcome of the gamble. Forty-three undergraduates judged the values of 230 distinct gambles, consisting of from 2 to 32 outcomes that were positive, zero, or negative, of different frequencies. The gambles were constructed from the union of four subdesigns.

The first subdesign used 55 gambles composed of two equally-likely outcomes, using all pairs of the following 11 values, -\$96, -\$72, -\$48, - \$24, -\$12, \$0, \$12, \$24, \$48, \$72, \$96. The second subdesign used 150 gambles containing from 2 to 32 equally likely outcomes of exactly two different values $(x, n_x; y, n_y)$; there were six pairs of values (x, y): $(-\$96, -\$48)$, $(-\$96, \$0)$, $(-\$96, \$96)$, $(-\$48, \$48)$, $(\$0, \$96)$, $(\$48, \$96)$; there were 5 different values of n_y ($n_y = 1, 2, 4, 8$, or 16 tickets) combined with 5 different values of n_x ($n_x = 1, 2, 4, 8$, or 16). The third subdesign contained 25 gambles of the form $(x, 1; \$96, n_y)$ with one ticket having one of 5 values of x $(-\$96, -\$48, \$0, \$24, \$48)$, factorially combined with 5 different values of n_y for $y = \$96$ ($n_y = 1, 2, 4, 8, 16$), producing probabilities, $n_x/(n_x + n_y)$, of .5, .667, .8, .889, or .941, to receive \$96. The fourth subdesign combined 5 different values of n_x for $x = \$0$, ($n_x = 1, 2, 4, 8, 16$) with 5 single values of $y = -\$96, -\$48, \$24, \$48, \$96$.

Birnbaum and McCormick fit the following model to the data:

$$U_I(x, n_x; y, n_y) = \frac{Au(x) + Bu(y)}{A + B} \tag{4a}$$

where $U_I(x, n_x; y, n_y)$ is the utility of the gamble from the receipt "indifference" point of view; $u(x)$ and $u(y)$ are the utilities of the outcomes; A and B are the weights of the outcomes, which depend on the number of outcomes of each value (n_x and n_y), the rank of the outcomes in the gamble (either lower or higher), and the augmented sign of the outcomes (the three levels of augmented sign, s_x, are +, 0, and −, for $x > 0$, $x = 0$, and $x < 0$, respectively):

$$A = W(n_x, r_x, s_x) \tag{4b}$$

$$B = W(n_y, r_y, s_y) \tag{4c}$$

where n_x and n_y are the number of outcomes; r_x and r_y are the ranks of the outcomes (i.e., either least or most in the gamble); s_x and s_y are the augmented signs of outcomes x and y, respectively. Because this experiment used five levels of number of outcomes, two levels of rank, and three levels of augmented sign, there are 30 weights to estimate (one of which can be fixed). Consistent with previous results (Birnbaum & Beeghley, 1997; Birnbaum et al., 1992; Birnbaum & McIntosh, 1996), the data could be as well fit with $u(x) = x$ as with a general power function, and the relationship between overt judgments and subjective values of Equation 4a could be approximated as linear.

The need to estimate the weight of the zero outcome separately from those for nonzero outcomes can be seen in Figure 7. Figure 7 shows mean judged indifference values for gambles with one ticket that is either \$0 (open circles) or \$24 (filled circles) and either 1, 2, 4, 8, or 16 tickets to win \$96 [$(\$0, 1; \$96, n_y)$ and $(\$24, 1; \$96, n_y)$]. Mean judgments are plotted as a function of the probability to win \$96 [i.e., $n_y/(1 + n_y)$]. Crossing of the curves in Figure 7 replicates the violation of monotonicity in Figure 1, using a yet another procedure for representing probability. The crossover in Figure 7 can be described by Expressions 4 if the weighting function depends on whether x is zero or positive.

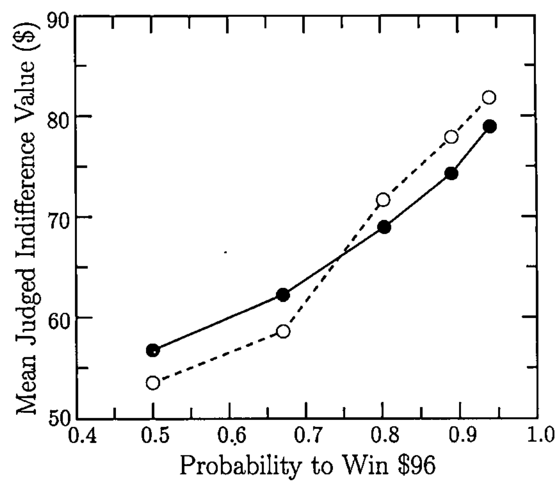

FIGURE 7. Mean judgment of receipt cash indifference value, as a function of probability to win \$96, with open and closed circles showing judgments when the lowest outcome was \$0 and \$24, respectively. Probabilities were manipulated by including one outcome of \$0 or \$24 with 1, 2, 4, 8, or 16 equally-likely outcomes of \$96. From Birnbaum and McCormick (1991).

The 30 weights of Expression 4 were estimated from the mean judgments of the 230 gambles used in the experiment. The estimated values of A (and B) in Equations 4b and 4c can be further simplified because they fit closely to the following multiplicative model:

$$W(n_x, r_x, s_x) = f(n_x)a_V(r_x, s_x) \tag{5}$$

where $f(n_x)$ is a function of number of outcomes, and $a_V(r_x, s_x)$ are six weights for two ranks and three augmented signs; these would be expected to depend on point of view, V. Fit of this model to the estimated weights can be assessed in Figure 8, which plots the estimated weights as a function of the estimated values of $f(n_x)$, with a separate curve for each level of rank and augmented sign. According to the multiplicative model of Equation 5, estimated weights should be linearly related to each other, with different slopes for different ranks and augmented signs, but they should share a common point of intersection. The estimated weights, shown as symbols, fall close to the bilinear fan predicted by the multiplicative model of Equation 5, shown as straight lines that intersect at a common point.

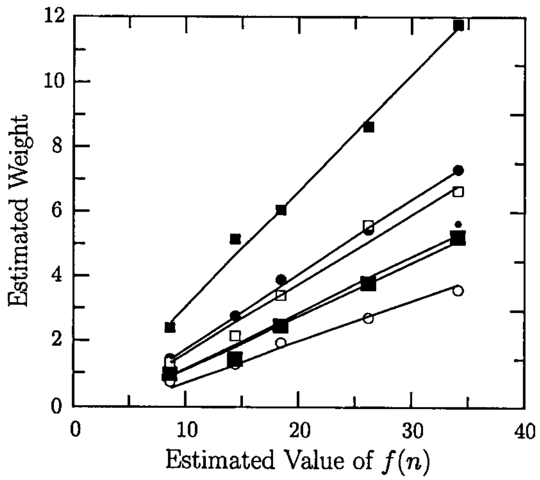

FIGURE 8. Estimated weights of $n = 1, 2, 4, 8$, or 16 equally likely outcomes, fit to judgments of 230 gambles in the "indifference" viewpoint, plotted as a function of the number of outcomes of a type, with a separate symbol for each combination of rank (circles = lower outcome, squares = higher outcome) and augmented sign (small = negative, unfilled = zero, large = positive) of outcomes. Lines show predictions of multiplicative model of $f(n)$ by a function of rank and augmented sign. From Birnbaum and McCormick (1991).

For rank = 1 (lower outcome), the estimated configural parameters, a_V in Equation 5, are .172, .127, and .232, for negative, zero, and positive outcomes, respectively. For rank = 2 (higher outcome), the parameters are .359, .216, and .165, respectively. Thus, for two positive outcomes, the relative weights of the lower and higher outcomes are .58 and .42, respectively, consistent with previous findings of "risk aversion" (greater weight on lower positive outcomes) in the neutral viewpoint. The smallest weight (.127) is for zero outcomes when zero is the lowest outcome; this reduced weight for zero accounts for violations of monotonicity. When the highest outcome is negative, it has greater weight (.359) than the lowest negative outcome (.172), consistent with previous findings of "risk seeking" for purely negative gambles (e.g., Tversky & Kahneman, 1992).

Equation 5 simplifies the treatment of monotonicity violations. Instead of two $S_x(p)$ functions, there is only one $f(n_x)$ function, the analog of $S(p)$, and the effects of rank and augmented sign are assumed to be multiplicative changes only, produced by different values of a_V.

If the a_V parameters were all equal, and if $f(n_x) = n_x$, then Equations 4 and 5 would reduce to expected utility theory. Instead, the estimated $f(n_x)$ function can be approximated as the square root of n_x. The fact that $f(n_x)$ follows this function implies that the relationship between relative weight $[A/(A + B)]$ and probability [averaged over different combinations with the same probability, $n_x/(n_x + n_y)$], will have an inverse-S relationship.

Varey, Mellers, and Birnbaum (1990) asked subjects to judge the proportion of dots of one color as a function of the numbers of dots of each color, and found a similar inverse-S relationship between average judged "proportion" and actual proportion. This function was explained by Varey et al. (1990) in terms of the psychophysical functions relating subjective number to actual number of dots in a relative ratio model. The psychophysical functions in that study were constrained to also account for judgments of "differences" and "ratios" of the numbers of dots (using subtractive and ratio models), as well as "proportions." A similar inverse-S weighting function has also been postulated by Tversky and Kahneman (1992) in their model of cumulative prospect theory, but it has a different interpretation in that theory. The difference in interpretations will be taken up in the next sections.

Cumulative Prospect Model of Binary Gambles. Tversky and Kahneman (1992) presented a special case of rank- and sign-dependent utility theory in which the weights of positive outcomes depend on the decumulative probability of each outcome in the gamble. For binary gambles of the form, $(z, 1 - p, x)$, where $0 \leq z < x$ and p is the probability to receive the higher outcome, the cumulative prospect model represents the value of the gamble as follows:

$$V(z, 1 - p, x) = (1 - \pi(x))u(z) + \pi(x)u(x) \tag{6a}$$

where $\pi(x) = W(p)$ is the weight of the higher outcome; the values of the two outcomes are $u(z)$ and $u(x)$.

Tversky and Kahneman (1992) approximated the value function with $u(x) = x^\beta$, where $\beta = .88$. In their model, the weight of the higher outcome, x, in a two outcome gamble is given by the following expression,

$$\pi(x) = W(p) = \frac{p^\gamma}{[p^\gamma + (1 - p)^\gamma]^{1/\gamma}} \tag{6b}$$

where the estimate of $\gamma = .61$ for positive outcomes. The certainty equivalent of a gamble is calculated from the inverse of the value function in Expression 6a,

$$\text{CE}(z, 1 - p, x) = V(z, 1 - p, x)^{1/\beta}. \tag{6c}$$

Tversky and Kahneman (1992) fit their model to transformed certainty equivalents, as shown in Figure 9. Each symbol represents a median certainty equivalent from Tversky and Kahneman (1992, Table 3), subtracting z, and divided by $x - z$. Unfilled squares, large circles, triangles, and small circles show adjusted certainty equivalents for $(\$0, p, x)$, where $x = \$50$, $\$100$, $\$200$, and $\$400$, respectively. The solid squares show results for gambles of the form $(\$50, 1-p, \$150)$. The solid curve in the figure shows Equation 6b, transformed by Equation 6c.

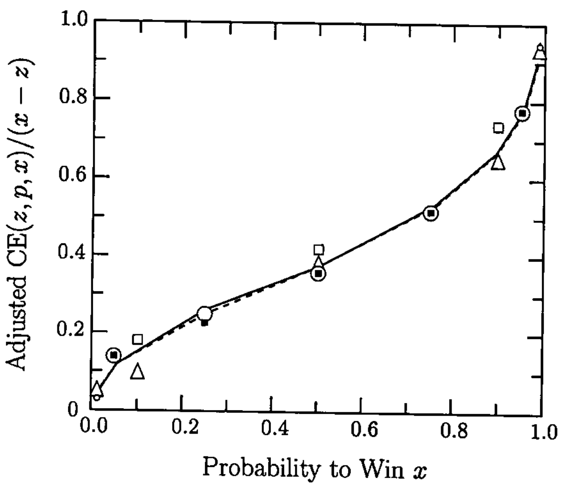

FIGURE 9. Median adjusted certainty equivalents of binary gambles plotted against the probability to win the larger outcome (Data from Tversky & Kahneman, 1992). Two weighting functions are compared. The solid line is the curve fit by Tversky and Kahneman (1992). The dashed curve shows that the configural weight model makes nearly identical predictions.

The experiment of Tversky and Kahneman (1992) was not designed to test for violations of monotonicity. Because of the confound between expected value and the set of comparisons offered for each gamble, and because of the use of relatively large lower positive outcomes ($50), their experiment is not well-suited for this purpose. However, if $u(x) = x$, for $\$0 \leq x \leq \$150$, if the solid squares fall on the same function, then there is no evidence of a different weighting function for zero outcomes for these data. If $u(x) = x^{.88}$, as suggested by Tversky and Kahneman, however, then the filled and unfilled circles in Figure 9 should not coincide unless the weighting function differs for these cases. That experiment cannot distinguish these interpretations.

The dashed curve plots the configural weight equations, Equations 4a and 5, assuming that the weight of the lower and higher outcomes are $A = a_V f(p)$ and $B = (1 - a_V)f(1 - p)$, where $f(p) = p^{.56}$ and $a_V = .63$ for the lower outcome. Equations 4a and 5 can be derived from the theory that subjects act as if they are minimizing an asymmetric squared loss function (weighted by a_V and $1 - a_V$ for squared overestimation and underestimation, respectively), by substituting $f(p)$ for p, and $f(1 - p)$ for $1 - p$ in Equation 6 of Birnbaum et al. (1992). The solid and dashed curves are virtually identical, but the implications are quite different, as becomes apparent for three-outcome gambles.

It is interesting that the value of $a_V = .63$, which fits the data of Tversky and Kahneman (1992) under the configural weighting interpretation, has also been obtained in two other experiments. It agrees with the weight of the lower positive outcome estimated in the neutral ("fair price") point of view by Birnbaum et al. (1992). The same value of a_V was also estimated from experiments testing interval independence (Birnbaum, Thompson, & Bean, in press). In one experiment, subjects judged the amount they would pay to receive one gamble rather than another. Subjects offered to pay an average of $44 to play ($74, $100) rather than ($8, $100) but they offered to pay only $24 to play ($6, $74) rather than ($6, $8). Ratings of strength of preference also showed that the judged strength of preference was greater when the common outcome was the highest than when it was the lowest outcome. Such violations of interval independence can be explained by greater weight on the lower outcome, and this experiment led to the value of .63 for that relative weight.

For two outcome gambles, the cumulative prospect model and the configural weight model (Equations 4a-4c and 5) make virtually identical predictions, as shown in Figure 9. However, for three outcome gambles, the theories make very different predictions for violations of branch independence, as will be shown in the next section.

When there are more than two outcomes, cumulative prospect theory postulates that the weights can be represented as differences in the $W(p)$ function for decumulative probability,

$$\pi(x_i) = W(P_i) - W(Q_i) \tag{7}$$

where $\pi(x_i)$ is the weight of outcome x_i in the gamble, P_i is the decumulative probability that the outcome in the gamble is $\geq x_i$; Q_i is the probability that the outcome exceeds x_i. For three, equally likely positive outcomes, the middle outcome would have the least weight, because $\pi(x) = W(2/3) - W(1/3)$ is the smallest of the three vertical differences (weights) in an inverse-S function, such as Figure 9.

7. VIOLATIONS OF BRANCH INDEPENDENCE

Savage's *sure thing principle* states that if two alternatives yield the same consequence for some state of the world, the value of that consequence should not make a difference for the preference due to other aspects of the alternatives. *Branch independence* corresponds to the "weak independence" condition that Cohen and Jaffray (1988) find more plausible than Savage's axiom. It states that if two gambles have the same outcome produced by the same event with a known probability, the value of that outcome should have no effect on the preference order. For three-outcome gambles $(x, p; y, q; z)$, in which the outcome is x with probability p, y with probability q and z otherwise $(1 - p - q)$, branch independence can be written as follows:

$$(x, p; y, q, z) \succ (x', p'; y', q'; z) \quad \text{if and only if} \quad (x, p; y, q; z') \succ (x', p'; y', q', z') \tag{8}$$

where $0 < p + q = p' + q' < 1$, and $\succ$ is the preference relation. Changing the common branch of z (with probability $1 - p - q$) to z' (at the same probability) should not affect the preference order produced by the other components of the gamble.

Branch independence is required by Savage's (1954) SEU theory, and is also implied by Edwards' (1954) psychological version of SEU that uses a weighting function of probability. It would also be observed if subjects were to edit the gambles being compared by canceling common branches, as discussed by Tversky (1969, 1972a), Kahneman and Tversky (1979), and Tversky and Kahneman (1992).

However, rank dependent utility theories allow violations of branch independence. For example, suppose that the rank-dependent utility of a gamble composed of three equally-likely outcomes (z, x, y), with outcomes chosen such that $0 < z < x' < x < y < y' < z'$, is given by the following expression,

$$\text{RDU}(z, x, y) = w_L u(z) + w_M u(x) + w_H u(y) \tag{9a}$$

where $\text{RDU}(z, x, y)$ is the rank-dependent utility of gamble (z, x, y); w_L, w_M, and w_H are the weights of the lowest, medium, and highest of three equally likely positive outcomes, respectively. When the common outcome is changed from lowest to highest (z to z'), then the weights of the low, medium, and highest outcomes are associated with x, y and z', respectively, as follows:

$$\text{RDU}(x, y, z') = w_L u(x) + w_M u(y) + w_H u(z'). \tag{9b}$$

Violations of Branch Independence in Choice. Birnbaum and McIntosh (1996) showed that Expressions 9a-9b imply that branch independence can be violated in two opposite ways. In the first case,

$$\frac{w_L}{w_M} < \frac{u(y') - u(y)}{u(x) - u(x')} < \frac{w_M}{w_H} \quad \text{if and only if} \quad (z, x, y) \succ (z, x', y') \text{ and } (x, y, z') \prec (x', y', z'). \tag{10}$$

In the other case,

$$\frac{w_L}{w_M} > \frac{u(y') - u(y)}{u(x) - u(x')} > \frac{w_M}{w_H} \quad \text{if and only if} \quad (z, x, y) \prec (z, x', y') \text{ and } (x, y, z') \succ (x', y', z'). \tag{11}$$

Experimentally, the tactic is to systematically vary x, y, x', and y' to find an intermediate value for the ratio of differences in utility, which should produce a reversal of preferences due to the change from z to z' (a violation of branch independence).

Birnbaum and McIntosh (1996) found systematic violations of branch independence in choices between gambles composed of three equally likely outcomes. They found that most people prefer ($2, $40, $44) to ($2, $10, $98); however, most people also prefer ($10, $98, $136) to ($40, $44, $136). This pattern was replicated with many different combinations of values. This pattern of preferences is consistent with Expression 10 but it is opposite that implied by Expression 11, which follows from the inverse-S weighting function of the cumulative prospect model of Tversky and Kahneman (1992).

According to the inverse-S (Equations 6b and 7), the middle of three equally likely outcomes should have the least weight. If the middle outcome had the least weight, then Expression 11 would follow because $w_L/w_M > 1$ and $1 > w_M/w_H$. These systematic violations are also not consistent with the theory that subjects edit and cancel common components when making choices, which implies that any violations of branch independence would be due to error (and should therefore be unsystematic).

Violations of Branch Independence in Judgment. Birnbaum and Beeghley (1997) found similar (but distinct) violations for buyer's and seller's prices. The violations of branch independence were again opposite those predicted by the inverse-S weighting function in both points of view. For example, ($4, $39, $45) was judged higher than ($4, $12, $96) yet ($39, $45, $148) was judged lower than ($12, $96, $148) in both viewpoints. However, judgments in the buyer's point of view of (x, y, z) decrease as $|x - y|$ is increased, holding $x + y$ constant, for all values of z; whereas, in the seller's point of view, these judgments increase as a function of $|x - y|$ when z is the highest outcome, but decrease when z is not highest. These changing violations and preference orders are consistent with the theory that the utility function of money is independent of the task, but that configural weights depend on the judge's point of view (Birnbaum et al., 1992; Birnbaum & Stegner, 1979; Birnbaum & Sutton, 1992).

Weights estimated from these three studies are presented in Table 1. In all three experiments, the different rank orders of the data could be well fit with the same utility function, $u(x) = x$ for $\$0 \leq x \leq \$150$. Although the weights differ in different tasks and viewpoints, all three sets of weights satisfy Expression 11. Weights from the choice task are intermediate between those obtained from judgments of buyer's prices and seller's prices, apparently closer to the buyer's viewpoint. The finding that all three experiments share the same utility function and weights that satisfy Expression 10 suggests that the pattern of violations is not due to something peculiar to either choice or judgment.

8. Testing Among Configural Weight Models

Until recently, most empirical work has been designed to distinguish configural weight theories from simpler, nonconfigural theories, rather than to test among alternative configural weight models. This section describes several properties that can be tested to distinguish among various models that have been suggested. These empirical properties are stochastic dominance, comonotonic independence, distribution independence, cumulative independence, and asymptotic independence.

Stochastic Dominance. Cumulative prospect theory implies stochastic dominance, whereas original prospect theory violates stochastic dominance (Kahneman

Experiment	Lowest	Middle	Highest
Buyer's Prices	.56	.36	.08
Seller's Prices	.27	.52	.21
Preferences	.51	.33	.16

TABLE 1. Estimated relative weights of three equally likely outcomes as a function of rank. Relative weights are normalized to sum to one by dividing by the sum of weights in each case. Weights for Preferences were estimated by Birnbaum and McIntosh (1996); Weights for Buyer's and Seller's Prices are from Birnbaum and Beeghley (in press). All three studies were fit with the same utility function $u(x) = x$ for $0 < x < \$150$.

& Tversky, 1979; Tversky & Kahneman, 1986, 1992). The configural weight model presented here violates monotonicity, therefore, it violates stochastic dominance. As shown below, this model also implies other violations of stochastic dominance that have not yet been tested.

According to the configural weight model presented here, for example, using the parameters of Birnbaum and McIntosh (1996), $U(\$12, .05; \$14, .05; \$96) = 53.6$, which is less than $U(\$12, .10; \$90, .05; \$96) = 61.1$, so the latter gamble should be judged better, even though the former stochastically dominates it. It seems worthwhile to test such predictions for violations of dominance using judged prices, using indirect comparisons in which each of the above gambles would be compared against a third gamble such as (\$55, .5; \$59, .5), and using direct comparisons between the two gambles. Note that this prediction of a violation of dominance does not rely on the presumed lower weighting for zero-valued outcomes, but follows instead from the configural weight model's weighting scheme for positive outcomes.

Although there have been occasional demonstrations of violations of stochastic dominance (e.g., Tversky & Kahneman, 1986), aside from the program of research reviewed here on violations of monotonicity, we do not yet have an adequate empirical description of more general types of violations of stochastic dominance.

Comonotonic Independence. Comonotonic independence is a special case of branch independence in which the preference order is assumed invariant when the common branch does not change rank order in the gambles to be compared.

For example, for three outcomes, comonotonic independence is the special case of Expression 8 where z and z' maintain the same rank in all four gambles (i.e., z and z' are either lowest in all four, middle in all four, or highest in all four gambles). A related property, ordinal (or "tail") independence, was tested by Wu (1994), who reported systematic violations that he attributed to a cancellation process specific to choice. Comonotonic independence has been tested in pure form (keeping the number of distinct outcomes equal in both gambles compared) in only a few papers (Birnbaum & Beeghley, 1997; Birnbaum & McIntosh, 1996; Wakker, Erev, & Weber, 1994; Weber & Kirsner, 1996), and it has not yet been reported to be systematically violated.

As noted by Birnbaum and McIntosh (1996), however, comonotonic independence has not yet received a strenuous test. Testing comonotonic independence evaluates the class of rank- and sign-dependent utility theories (Luce & Fishburn, 1991, 1995). This class includes cumulative prospect theory (Tversky & Kahneman, 1992) and the model presented here, both of which satisfy comonotonic independence when the probability distribution is fixed.

Chew and Wakker (1996) discuss the comonotonic sure thing principle as characterizing "all existing rank-dependent forms," but it is important to note that their treatment does not include all configural forms. If configural weights depend on the spacing of the outcomes as well as their ranks, as they would according to the minimum loss theory presented by Birnbaum et al. (1992), then comonotonic independence can be violated (Birnbaum & McIntosh, 1996, Appendix A).

Distribution Independence. Distribution independence assumes that preferences should be independent of the (common) probability distribution of common branches. For four outcome gambles, with outcomes chosen such that $0 < z < x' < x < y < y' < v$, and nonzero probabilities, p, q, r, and $s = 1-p-q-r$, distribution independence requires:

$$\begin{gathered}(z, r; x, p; y, q; v, s) \succ (z, r; x', p; y', q; v, s)\\ \text{if and only if}\\ (z, s; x, p; y, q; v, r) \succ (z, s; x', p; y', q; v, r)\end{gathered} \tag{12}$$

Distribution independence asserts that the trade-off between $(x, p; y, q)$ and $(x', p; y', q)$ should be independent of the probability distribution of the common branches (r and s vs. s and r), holding (p, q) fixed. Note that in Expression 12, the common outcomes are the same, but their probabilities differ; whereas, in branch independence the probabilities of the common branches are the same and their outcomes differ.

The configural weight model presented in this chapter can violate branch independence but must satisfy distribution independence. The revised configural weight model of Birnbaum and Stegner (1979, Equation 10), however, violates distribution independence. This model will be discussed further in the section below on asymptotic independence.

Cumulative prospect theory implies systematic violations of distribution independence, with the pattern of violations dependent on the $W(p)$ function of Equation 7. For example, the model of Tversky and Kahneman (1992) implies that

$$(\$2, .59; \$10, .2; \$98, .2; \$108, .01) \succ (\$2, .59; \$50, .2; \$54, .2; \$108, .01);$$

however,

$$(\$2, .01; \$10, .2; \$98, .2; \$108, .59) \prec (\$2, .01; \$50, .2; \$54, .2; \$108, .59),$$

violating distribution independence.

Birnbaum and Chavez (1996) tested distribution independence, and found small but systematic violations in the opposite direction from those predicted by cumulative prospect model. For example, they found that the percentage choosing $(z, .59; x, .2; y, .2; v, .01)$ over $(z, .59; x', .2; y', .2; v, .01)$ is greater than the percentage choosing $(z, .01; x, .2; y, .2; v, .59)$ over $(z, .01; x', .2; y', .2; v, .59)$ for all six different contrasts of (x,y) vs. (x',y') used, contrary to the prediction of the inverse-S function, which predicts the opposite pattern of shifting preferences. Similar results were obtained when $(r,s) = (.55, .05)$.

Cumulative Independence. If the weights depend entirely on the cumulative (or decumulative) distribution of outcomes, as in Equation 7, then the weights of outcomes should be independent of how that cumulative (decumulative) distribution is produced. Cumulative independence holds for cumulative prospect theory and is systematically violated by configural weight models.

Cumulative prospect theory makes the following predictions for two and three-outcome gambles, with nonzero probabilities, p, q, and r that sum to one, and $0 < z < x' < x < y < y' < z'$:

$$\text{If } (z,r;x,p;y,q) \succ (z,r;x',p;y',q), \text{ then } (x',r;y,p+q) \succ (x',r+p;y',q). \quad (13)$$

Similarly,

$$\text{If } (x',p;y',q;z',r) \succ (x,p;y,q;z',r), \text{ then } (x',p;y',q+r) \succ (x,p+q;y',r). \quad (14)$$

These tests of cumulative independence do not assume a particular form of $W(p)$, such as the inverse-S, but hold for any cumulative (or decumulative) weighting function. These tests are not "pure" tests of a single axiom, as they can be viewed as a combination of comonotonic independence, monotonicity, transitivity, and the "accounting equivalence" that equal outcomes can be coalesced by adding their probabilities. [For example, one can deduce Expression 13 as follows: If $(z,r;x,p;y,q) \succ (z,r;x',p;y',q)$ then $(x',r;x,p;y,q) \succ (x',r;x',p;y',q)$ by comonotonic independence; by monotonicity, $(x',r;y,p;y,q) \succ (x',r;x,p;y,q)$; therefore, by transitivity, $(x',r;y,p;y,q) \succ (x',r;x',p;y',q)$; finally, by the coalescing equivalence, $(x',r;y,p+q) \succ (x',r+p;y',q)$.] The key idea of cumulative independence is that increasing the probability of an outcome should have the same effect on weights as adding another distinct outcome with the same probability, if that outcome preserves comonotonicity.

Configural weight theory, on the other hand, distinguishes increasing the probability of an outcome from adding a new outcome. If $S(p)$ is negatively accelerated, then a new outcome will have greater weight than the marginal increase in weight due to the same increase in probability of an existing outcome. The configural weight model presented here implies violations of the cumulative independence conditions described above. For example, with $p = q = r = 1/3$, using parameters from Birnbaum and McIntosh (1996), the model implies the following violations of cumulative independence:

$$U(\$2, 1/3; \$40, 1/3; \$44, 1/3) = 21.3 > U(\$2, 1/3; \$10, 1/3; \$98, 1/3) = 20.0;$$

however,

$$U(\$10, 1/3; \$44, 2/3) = 25.9 < U(\$10, 2/3; \$98, 1/3) = 35.4,$$

in contradiction to Expression 13.

Similarly,

$$U(\$10, 1/3; \$98, 1/3; \$108, 1/3) = 55.7 > U(\$40, 1/3; \$44, 1/3; \$108, 1/3) = 52.2;$$

however,

$$U(\$10, 1/3; \$98, 2/3) = 51.2 < U(\$40, 2/3; \$98, 1/3) = 56.7,$$

contradicting Expression 14.

This property appears to give a sharp distinction between cumulative prospect theory and the configural weight model presented here. Although the property has not yet been tested in a single experiment with the same subjects, data by Wu and Gonzalez (1996) combined with data of Birnbaum and McIntosh (1996) suggest indirectly that the property might be violated.

Asymptotic Independence. Birnbaum and Stegner (1979, Equation 10) presented a revised configural weight model in which the transfer of weights among outcomes of different ranks depends on the point of view of the judge and is also proportional to the weight of the outcome losing weight. This revised model differs from the previous rank-dependent configural weight model of Birnbaum (1974), extended by Birnbaum et al. (1992), and presented here. The revised configural weight model of Birnbaum and Stegner (1979) violates both distribution independence and asymptotic independence.

For two outcome gambles, asymptotic lower independence can be defined as follows: as $p \to 0$, $U(x,p,y) \to u(y)$, for all x. Asymptotic upper independence is defined as follows: as $1-p \to 0$, $U(x,p,y) \to u(x)$, for all y. Thus, the value of an improbable outcome should become less and less relevant as the probability of the other outcome approaches 1.

For moral judgment (Birnbaum, 1973; Riskey & Birnbaum, 1974), likeableness judgments (Birnbaum & Rose, 1973), and buying prices (Birnbaum & Stegner, 1979), however, the value of the worst deed, trait, or estimate appears to set an upper limit on a person's morality, likeableness, or buying price. Given a person has done a single very immoral deed, for example, it appears that the person's judged morality is bounded to be low, no matter how many good deeds that person does. However, a single good deed appears to set no such limit on the lower bound of judged morality.

For gambles, asymptotic independence says that no matter how bad an outcome is, it should approach irrelevance as it becomes less and less probable. A contrary notion, for example, is that some outcomes are so bad that no matter how small their probabilities, the utility of a gamble with such a possible outcome is bounded to a lower value unless its probability is zero. Discussions of insurance and risk of accidental nuclear war, for example, often seem to express this notion. The aversion that people have toward probabilistic insurance (a less than half-priced policy in which the insurance agent flips a coin to decide if the company will pay off in the event of a fire) suggests that asymptotic independence may be violated. People often express the idea that the purpose of insurance is to eliminate the possibility of bad outcomes, rather than to merely reduce their probabilities.

The revised configural weight model presented by Birnbaum and Stegner (1979, Equation 10) allows an outcome of near zero probability to place an upper (or lower) bound on the response as the probability of that outcome approaches (but does not equal) zero. For buying prices of two outcome gambles, (x,p,y), $x < y$, $0 < p < 1$, assuming that the lower outcome receives greater weight, this revised model retains Equation 1a, but it replaces Equations 1b and 1c with the following:

$$A = S_x(p) + a_V S_x(1-p) \tag{15a}$$

$$B = (1-a_V)S_x(1-p) \tag{15b}$$

where A and B are the absolute configural weights of the lower and higher outcomes, respectively, as in Equation 1a; and a_V is the configural weight parameter that in this model represents the proportion of weight taken from the higher valued outcome (for buying prices) and given to the lower valued outcome. The other terms are as defined in Expressions 1.

For example, if $S(p) = p$, and $a_V > 0$, Equations 15 imply that $U(x,p,y) = [p + a_V(1-p)]u(x) + [(1-a_V)(1-p)]u(y)$. As $p \to 0$, $U(x,p,y) \to a_V u(x) + (1-a_V)u(y)$, which indicates that as long as the lower outcome is possible, it limits the utility of the gamble. However, as $1-p \to 0$, $U(x,p,y) \to u(x)$. Thus, this model violates asymptotic lower independence, but satisfies asymptotic upper independence.

When weight is transferred from the lower to the higher value, as for example in selling prices, then the weight transferred is proportional to the weight of the lower value, as follows:

$$A = (1-a_V)S_x(p) \tag{16a}$$

$$B = S_x(1-p) + a_V S_x(p). \tag{16b}$$

In this case, a possible good outcome sets a lower limit on the selling price, but a low outcome sets no such upper limit.

The models in Equations 15-16 violate asymptotic independence, implying that the worst outcome places an upper bound on the buying price and the best outcome sets a lower bound on the selling price of a gamble. This revised model gave a better fit (than the simple configural model) to judgments of buying and selling prices of used cars based on estimates given by sources (Birnbaum & Stegner, 1979), and it can describe judgments of likeableness and morality. However, implications of asymptotic independence have not been tested for judgments or choices among gambles.

9. Discussion and Conclusions

Violations of monotonicity add to a growing literature in judgment and decision making of phenomena that trouble the theoretician. Taking the results from different studies together, what might be considered a single empirical effect, the pattern of results in Figure 1, might show up as violations of three axioms: monotonicity, transitivity, and consistency.

Birnbaum and Sutton (1992) noted that because the monotonicity violation is obtained in judgment but not direct choice, there is a reversal of preference, violating consistency. Let $A = (\$0, .05, \$96)$ and $B = (\$24, .05, \$96)$. Birnbaum and Sutton found that A is judged higher than B, but in a direct comparison, the vast majority choose B over A.

Birnbaum and Thompson (1996) found evidence suggesting that there is a value of cash that is intermediate between A and B, such that $P(A, c) > 1/2$ and $P(c, B) > 1/2$. However, from Birnbaum and Sutton, $P(B, A) > 1/2$, which might be taken as a violation of transitivity. It is unclear if such cross-experiment comparisons predict what a single individual would do when faced with all three comparisons, but it should be clear that the theoretician has a problem accounting for all of the choices in terms of a single, transitive preference order.

The axioms of monotonicity, transitivity, and consistency appear quite reasonable from a normative standpoint. Luce and von Winterfeldt (1994) regard transitivity as "nonnegotiable" from the normative perspective. In judgment experiments, where the subject assigns a number to each gamble, transitivity is automatically satisfied (because the numbers are transitive). However, Tversky (1969) concluded that there are situations in which transitivity is systematically violated in choice. If choices can be made to violate transitivity, therefore, one might argue that judgment should be preferred as a mode of response because it satisfies transitivity.

On the other hand, this chapter reviews evidence that monotonicity can be violated in judgment, but has not been violated systematically in transparent choices. Because monotonicity is an axiom that seems compelling to both theoreticians and subjects, who rarely try to defend their violations, choice might seem a preferred method because it seems to obey the axiom of monotonicity.

Thus, if we try to enforce the most cherished of normative axioms by our selection of procedure, we are torn between choice, which presumably satisfies monotonicity but might violate transitivity, and judgment, which automatically satisfies transitivity but may systematically violate monotonicity.

The intermediate method of choice-based certainty equivalents might therefore seem a good compromise between choice and judgment. Certainty equivalents satisfy transitivity. Some evidence suggests that certainty equivalents based on PEST may reduce violations of monotonicity (von Winterfeldt et al., 1997). However, violations of scalability (Birnbaum & Thompson, 1996) indicate that violations of monotonicity depend on the value of cash against which the gambles are compared. Furthermore, contextual effects in choices (Birnbaum, 1992b) suggest that we need more data and better theory on this procedure before we can know how to distinguish the value of a gamble from the context of cash values presented in the procedure.

The classic form of preference reversals (Bostic et al., 1990; Lichtenstein & Slovic, 1971; Lindman, 1971; Slovic, Lichtenstein, & Fischhoff, 1988) are but a small portion of the reversals of preference that have now been demonstrated. The problems at hand are to explain how and why the apparent rank order of gambles changes depending on the task (choice vs. judgment), the judge's point of view (buyer's vs. neutral's, vs. seller's), the common outcome (branch independence), the context, and whether outcomes are negative, positive, or zero.

Because so many factors appear to affect preferences, Tversky and Kahneman (1992) concluded with the "pessimistic" assessment that no decision theory will successfully account for all of the phenomena. Indeed, this chapter has reviewed results that go beyond even the list of problems discussed by Tversky and Kahneman. Nevertheless, analogies from the history of science give us room for hope of devising a single theory that can account for all of the phenomena.

What can at first appear to be many exceptions and complications in one theory can suddenly fall into place when a better theory is devised. For example, planetary positions calculated from Ptolemy's geocentric model with uniform circular motion required many "fudge" factors of offset epicycles to fit the data. The heliocentric model of Copernicus was simpler, but still required "fudge" factors to account for departures from uniform circular motion. Kepler's elliptical models (Kepler's "laws") produced a more accurate description with a simpler unifying set of equations.

In decision research, changes in preference order due to the subject's viewpoint, for example, would be interpreted as evidence of changing $u(x)$ functions in the framework of nonconfigural models. However, configural weight models allow one to retain the premise that the $u(x)$ function is invariant with respect to viewpoint (Birnbaum et al., 1992; Birnbaum & Sutton, 1992). Furthermore, configural weight models can account for violations of branch independence in different viewpoints, again using a single $u(x)$ function. Evidence so far does not yet require the rejection of a single $S(p)$ function, if configural weights are allowed to depend on the number of outcomes, their augmented signs, and ranks.

The configural weight models, on the other hand, contain these configural weighting parameters, which until they can be explained by deeper primitives, seem to have the character of the epicycles used early in Astronomy. Different ideas about the origin of the configural weights – that they depend on asymmetric costs of over- or under-estimation (Birnbaum et al., 1992; Birnbaum & McIntosh, 1996; Weber, 1994), that they depend on subject's conformance to the comonotonic "sure thing" principle (e.g., Chew & Wakker, 1996), or properties of joint receipt (Luce, 1995b; Cho et al., 1994) – lead to distinct testable implications. These implications, some of which are described in Section 8, have the potential to make the world seem even more complicated, and hopefully, they may also lead to new theory that will make it seem simpler.