Journal of Risk and Uncertainty, 2: 61-104 (1989)

An Experimental Test of Several Generalized Utility Theories

COLIN F. CAMERER
The Wharton School, University of Pennsylvania

Key words: expected utility, generalized utility, nonexpected utility, risky choice, Allais paradox

Abstract

There is much evidence that people willingly violate expected utility theory when making choices. Several axiomatic theories have been proposed to explain some of this evidence, but there are few data that discriminate between the theories. To gather such data, an experiment was conducted using pairs of gambles with three levels of outcomes and many combinations of probabilities. Most typical findings were replicated, including the common consequence effect and different risk attitudes for gains and losses. There is evidence of both fanning out and fanning in of indifference curves, and both quasiconcavity and quasiconvexity of preferences. No theory can explain all the data, but prospect theory and the hypothesis that indifference curves fan out can explain most of them.

Expected utility theory (EU) is the foundation of the economics of uncertainty and the focus of much research and application in decision theory and psychology (Schoemaker, 1982). EU rests on the proof that a utility function exists that makes preferences over gambles representable by the numerical expected utilities of the gambles—preferred gambles have higher numbers—if preferences obey several simple axioms. Despite the intuitive appeal of the axioms, many patterns of choices that violate them have been pointed out (most notably by Allais, 1953, 1979; Ellsberg, 1961; and Kahneman and Tversky, 1979).

Subjects who violate EU axioms in their choices often change their preferences to conform to the axioms when their violations are pointed out. But many well-informed subjects refuse to change their preferences; they reject the axioms instead. Animals violate EU in the same way people do (Battalio, Kagel, and MacDonald, 1985), which suggests the violations have a simple cause, perhaps perceptual.

To describe the choices such people make, theories have been developed in which the EU axioms are weakened to allow patterns of preference that violate EU. Except for prospect theory, all these new theories include EU as a special case, so they will be called *generalized utility theories*. Many theories will not be considered.[1] The theories (and authors) we will consider are weighted utility theory (Chew and MacCrimmon) and its kin, skew-symmetric bilinear (SSB) utility theory (Fishburn) and regret theory (Bell, Loomes and Sugden); implicit expected utility theory (Chew, Dekel); the fanning-out hypothesis (Machina); lottery-dependent expected utility (Becker and Sarin); prospect theory (Kahneman and Tversky); and expected utility with rank-dependent probabilities (Quiggin, Yaari).

Since most of these theories were developed to explain the same small set of EU violations, when they were created there were few data which distinguish among them empirically. Some new data do distinguish some of the theories from each other. These data will be reviewed below, along with a new test that can distinguish all the theories.

We will not discuss many EU violations that are especially difficult to explain. These include violations of transitivity, perhaps resulting from reversals of preference or response-mode effects (Grether and Plott, 1979; Slovic and Lichtenstein, 1983; Hershey, Kunreuther, and Schoemaker, 1982); aversion to vagueness or ambiguity (Ellsberg, 1961; Einhorn and Hogarth, 1985; Hogarth and Kunreuther, forthcoming); violations of stochastic dominance due to opacity (Tversky and Kahneman, 1987); and the effects of framing, such as reference-point shifts (Kahneman and Tversky, 1979) and statistical correlation of lotteries (Loomes and Sugden, 1987a,b; Loomes, 1988).

1. Theoretical predictions

Each generalized utility theory we consider makes a slightly different prediction about the shape of curves of indifference between sets of gambles. By graphing predicted indifference curves and studying their properties, it is apparent what choices between gambles can separate all the theories. A set of such gambles are then offered to subjects;[2] their choices test the theories.

Predictions about indifference curves are easily displayed in the triangle diagram developed by Marschak (1950) and put to good use by Machina (1982). (Indeed, this section overlaps heavily with Machina, 1982, 1983; Sugden, 1986; and Weber and Camerer, 1987.) Consider three gambles X_L, X_M, X_H (low, medium, high) with objective probabilities p_L, p_M, p_H, such that $X_L \prec X_M$ (X_M is preferred to X_L) and $X_M \prec X_H$. (Indifference is denoted $X_i \sim X_j$.) Outcomes are represented by degenerate lotteries that give a certain result with probability one. (In some cases we also denote outcomes by lower-case letters.) The gambles X_L, X_M, X_H our subjects won were degenerate lotteries with certain monetary outcomes.

If three gambles are fixed, then the set of all gambles over those gambles can be represented in two dimensions, in $p_L - p_H$ space. The third dimension, p_M, is implicit in the graph because $p_M = 1 - p_L - p_H$. Since the sum of the probabilities cannot be greater than one, the set of feasible probabilities is a triangle bounded by the lines $p_L = 0$ (the left edge), $p_M = 0$ (the hypotenuse), and $p_H = 0$ (the lower edge), as shown in some of the figures below.

1.1. Expected utility theory

In its simplest form, EU requires three axioms: ordering, continuity, and independence. In describing these axioms, and below, we assume that gambles are probability distributions on final wealth states.

Ordering requires that preferences be a *weak order* (Fishburn, 1970): asymmetric, and negatively transitive. These simple properties imply stronger ones, including completeness and transitivity. Completeness means that either $X \prec Y$, $X \succ Y$, or $X \sim Y$ (i.e., people can make up their minds). Transitivity means that $X \prec Y$ and $Y \prec Z$ implies $X \prec Z$. In the triangle diagram, completeness implies that two points either lie on different indifference curves or lie on the same curve. Transitivity implies that indifference curves do not cross inside the triangle.

Continuity requires that for any gamble G_m that satisfies $G_l \prec G_m \prec G_h$, a unique probability q can be found for which one is indifferent between G_m and a gamble with a q chance of G_h and a $1 - q$ chance of G_l (written $G_m \sim qG_h + (1 - q)G_l$). Continuity ensures that every gamble lies on some indifference curve; there are no holes in the indifference map in the triangle. Uniqueness of the probability q implies that indifference curves are not thick.

Independence assumes that if the gambles X and Y are equally preferable, then the gambles composed of a q chance of X (or Y) and a $1 - q$ chance of Z are also equally preferable. That is, $X \sim Y$ implies $qX + (1 - q)Z \sim qY + (1 - q)Z$ for any q and Z. The independence axiom implies that indifference curves are parallel straight lines (see Marschak, 1950, and below).

Variants of ordering and continuity are required in virtually all axiomatized theories of choices. Empirical criticism of them is relatively rare and unconvincing (see review by MacCrimmon and Larsson, 1979[3]). The independence axiom is the source of many of the violations of EU mentioned above; generalized theories typically weaken it.

The ordering axiom implies that a numerical scale exists that represents a person's preferences over gambles (more-preferred gambles have higher numbers). Continuity and independence make the numerical utility of a gamble equal the expectation of the utilities of the gamble's possible outcomes. That is,

$$\text{EU:} \qquad U(qX + (1 - q)Y) = qU(X) + (1 - q)U(Y). \qquad (1)$$

Furthermore, any positive affine transformation $V(X)$ of the utility function $U(X)$ (i.e., $V(X) = a + bU(X)$, $b > 0$) will represent preferences equivalently.[4]

In EU, an indifference curve in the triangle diagram is a set of gambles with the same expected utility U^*,

$$\text{EU:} \qquad U^* = p_L U(X_L) + p_M U(X_M) + p_H U(X_H). \qquad (2)$$

Substituting $p_M = 1 - p_L - p_H$ and rewriting (2) in slope-intercept form,

$$p_H = \frac{U^* - U(X_M)}{U(X_H) - U(X_M)} + p_L \frac{U(X_M) - U(X_L)}{U(X_H) - U(X_M)}. \quad (3)$$

The slope of the tangent line to an indifference curve at a point is dp_H/dp_L:

$$\frac{dp_H}{dp_L} = \frac{U(X_M) - U(X_L)}{U(X_H) - U_{(X_M)}}. \quad (4)$$

Since the slope dp_H/dp_L is a constant, depending only on the relative utilities of the three outcomes, the indifference curves are straight lines with the same slope, as shown in figure 1. The properties of indifference curves under EU, linearity and parallelism, are shown in table 1 and contrasted with the properties of curves under other theories, derived below. (Casual readers can look at table 1 and then skip ahead to section 3.)

The slope $[U(X_M) - U(X_L)]/[U(X_H) - U(X_M)]$ can be naturally interpreted as a discrete marginal rate of substitution of p_H for p_L, or as the shadow price of probabilistic units of the highly valued gamble in terms of probabilistic units of the lowest-valued gamble. Risk-averse people will have a larger $U(X_M)$ (relative to

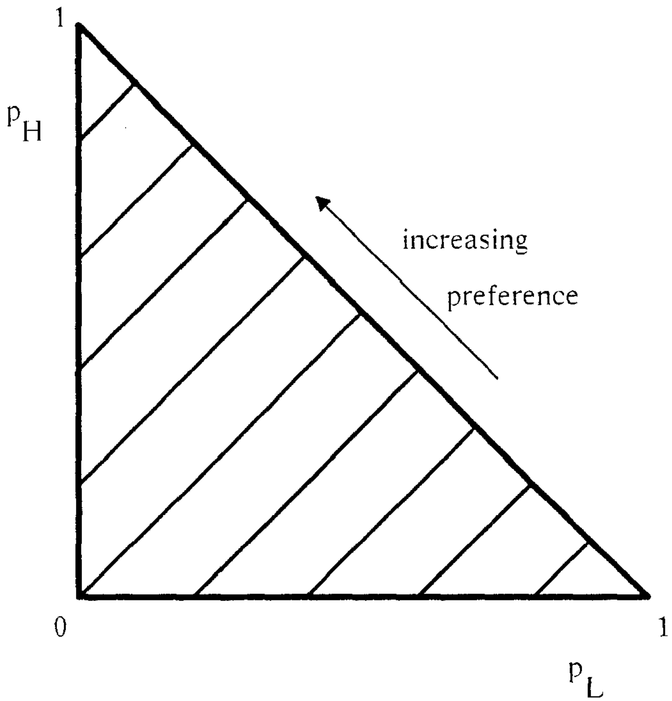

Fig. 1. Indifference curves assuming expected utility.

Table 1. Predictions of competing theories about properties of indifference curves

		Properties of curves			
Theory (figure no.)	Functional form for $U^*(F) = U(qX + (1-q)Y(X<Y)$	Straight lines?	Fanning out?	Fanning in?	Miscellaneous
Expected utility (1)	$qU(X) + (1-q)U(Y)$	Yes	No	No	Curves parallel
Weighted utility (2)	$\dfrac{qW(X)U(X) + (1-q)W(Y)U(Y)}{qW(X) + (1-q)W(Y)}$ $W(X) < 1$	Yes	Yes	No	Curves meet in a point
	$W(X) > 1$	Yes	No	Yes	
Implicit expected utility (3)	$qU(X,U^*) + (1-q)U(Y,U^*)$	Yes	Maybe	Maybe	Only testable property is betweenn[illegible]
The fanning-out hypothesis (4)	$\dfrac{-U''(x;F)}{U'(x;F)} \geqslant \dfrac{-U''(x,G)}{U'(x;G)}$ if $F(x) \leqslant G(x)$ for all x	Maybe	Yes	No	Movements to northwe[illegible] cause steeper slopes
Lottery-dependent utility (5)	$qU(X,C_F) + (1-q)U(Y,C_F)$ $h(x)$ concave	No	Yes	No	Curves concave
	$C_F = \int h(X)dF(X)$ $h(x)$ convex	No	Yes	No	Curves convex
Prospect theory (6)	$u(X) + \pi(q)U(Y-X)$ $\pi(q)U(X) + \pi(r)U(Y), q + r < 1$	No	Lower edge	Left edge, hypotenuse	Parallel along $P_H = (1 - P_L)/2$
Rank-dependent utility (7)	$g(q)U(X) + (1 - g(q))U(Y)$ g concave	No	Lower edge	Left edge	Parallel along hypotenuse
	g convex	No	Left edge	Lower edge	

$U(X_L)$ and $U(X_H)$); their indifference curves are steeper; and they demand a higher price to bear risk. The independence axiom forces the degree of risk-aversion at a particular point to be independent of the location of that point in the triangle.

Machina (1982) pointed out that the slope $[U(X_M) - U(X_L)]/[U(X_H) - U(X_M)]$ can be rewritten as

$$1 - \frac{\{[U(X_H) - U(X_M)] - [U(X_M) - U(X_L)]\}}{U(X_H) - U(X_M)}. \tag{5}$$

The second term in (5) is the discrete analogue of the familiar Arrow-Pratt measure of risk aversion, $-U''(X)/U'(X)$, because the numerator represents a difference of differences, and the denominator represents a difference. Thus, the slope of indifference curves (plus one) is a measure of risk aversion much like the Arrow-Pratt measure.

1.2. Weighted utility theory

Chew and MacCrimmon (1979) used a weakened version of the EU independence axaiom to derive weighted utility theory (see also Chew, 1983). Weak independence assumes that for each probability q, there is a probability r for which $X \sim Y$ implies $qX + (1 - q)Z \sim rY + (1 - r)Z$ for any Z. (EU independence, sometimes called *strong independence,* requires $r = q$.) People are effectively allowed an extra degree of freedom in choosing the probability r (depending on q) that satisfies weak independence for any Z. The extra degree of freedom is manifested in a novel weighting function that combines with probabilities and utilities. Preferences can be represented by a weighted expected utility,

$$\text{WEU:} \qquad U(qX + (1 - q)Y) = \frac{qW(X)U(X) + (1 - q)W(Y)U(Y)}{qW(X) + (1 - q)W(Y)}. \tag{6}$$

The weighting function $W(X)$ is, roughly speaking, determined up to a positive multiplicative constant (see Chew, 1983, p. 1076). If $W(\cdot)$ is constant, the weights in (6) divide out and weighted utility reduces to EU.

The axioms suggest no obvious psychological interpretation to the weighting function. The weights seem to modify probabilities, reflecting mental distortions or misperceptions, to a degree that depends on outcomes. Weber (1982) interpreted $W(X)$ as the conceivability or vividity of X.[5]

In our simple three-outcome setting, weighted utility predicts indifference curves of the form

$$p_H = \frac{W(X_M)(U^* - U(X_M)) + p_L[W(X_L)(U^* - U(X_L)) + W(X_M)(U(X_M) - U^*)]}{W(X_M)(U^* - U(X_M)) + W(X_H)(U(X_H) - U^*)} \tag{7}$$

with tangent line slopes of

$$\frac{dp_H}{dp_L} = \frac{W(X_L)(U^* - U(X_L)) + W(X_M)(U(X_M) - U^*)}{W(X_M)(U^* - U(X_M)) + W(X_H)(U(X_H) - U^*)}. \tag{8}$$

The tangent line slopes in (8) depend on U^*, so each indifference curve has a different slope (equation (7) shows that the curves are straight lines). If we arbitrarily set $W(X_L) = W(X_H) = 1$ (we have the freedom to do so), then either $W(X_M) < 1$ or $W(X_M) > 1$. If $W(X_M) < 1$ (called the light hypothesis by Chew and Waller, 1986) then dp_H/dp_L is increasing in U^*; curves get steeper from right to left. $W(X_M) > 1$ (the heavy hypothesis) implies slopes are decreasing in U^*; curves get flatter. It is easy to show that the indifference curves all meet in a point outside the triangle, as shown in figure 2 (for $W(X_M) < 1$). (In EU the meeting point is infinitely far away, so the curves are parallel.) Curves that get uniformly steeper as one goes from right to left ($W(X_M) < 1$) are said to *fan out* (Machina, 1982); curves that get flatter *fan in*.

Weighted utility has a remarkable kinship with the *skew-symmetric bilinear* (SSB) utility studied by Fishburn (1982, 1983, 1984). In SSB utility, preferences are defined over *pairs* of gambles (X, Y) and represented by a numerical function

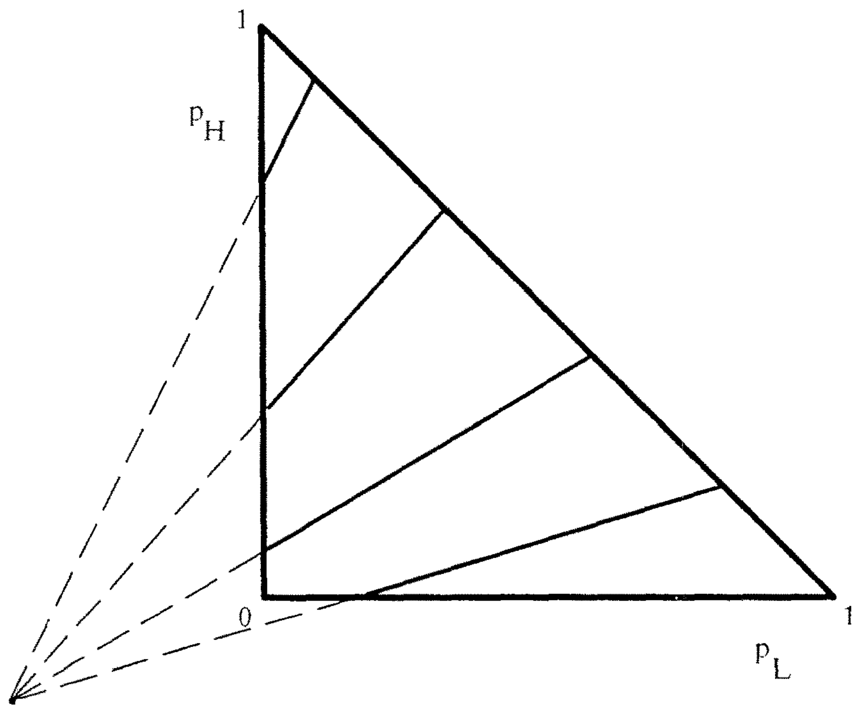

Fig. 2. Indifference curves assuming weighted utility ($W(X_M) < 1$).

$\phi(X,Y)$. If $X \prec Y$, $\phi(X,Y) < 0$; if $X \sim Y$, $\phi(X,Y) = 0$; and if $Y \prec X$, $\phi(X,Y) > 0$. (If $\phi(X,Y) = U(X) - U(Y)$, then SSB reduces to EU.) The function $\phi(X,Y)$ is skew-symmetric if $\phi(X,Y) = -\phi(X,Y)$ and bilinear if it is linear in each argument (e.g., $\phi(qX + (1-q)Z,Y) = q\phi(X,Y) + (1-q)\phi(Z,Y)$). If X and Y are uncorrelated lotteries with component outcomes x_i and y_i with probabilities p_i and q_i, then $\phi(X,Y) = \Sigma_{i=1}^{n} p_i q_i \phi(x_i,y_i)$.

Fishburn proved that an SSB function $\phi(X,Y)$ represents preferences if they satisfy a symmetry axiom: If $X_M \sim .5X_L + .5X_H$, p_1 must equal p_2 in the indifference relations $.5X_M + .5X_H \sim p_1X_L + (1-p_1)X_H$ and $.5X_M + .5X_L \sim p_2X_H + (1-p_2)X_L$. (In EU, $p_1 = p_2 = .25$.) That is, indifference curves in a triangle diagram with $X_M \sim .5X_L + .5X_H$ must be symmetric around the 45-degree line. (Curves must be straight lines because of a dominance axiom that ensures bilinearity.) The symmetry axiom allows the slopes of indifference curves to vary, but they must vary symmetrically so they will meet at a common point outside the triangle, as in weighted utility.

If transitivity is assumed, then SSB utility is equivalent to weighted utility. This kinship is rather remarkable because weighted utility was generated by weakening the independence axiom to explain the Allais paradox, while SSB demonstrates that a utility function on pairs of gambles can be useful. Of course, if $\phi(X,Y) = V(X)W(Y) - V(Y)W(X)$ (with $V(X) = W(X)U(X)$) then the equivalence of SSB and weighted utility is rather clear. If transitivity is not assumed, then SSB indifference curves can meet at a point *inside* the triangle; intransitivity results from moving from curve to curve in a counterclockwise cycle (see figure 9 in Machina, 1987).

Regret theory is an important generalization of intransitive SSB that applies when outcomes of different lotteries may be statistically correlated (Bell, 1982, 1983; Loomes and Sugden, 1982, 1987a). (Since transitive SSB is a special case, regret theory need not imply intransitivities, but they are permitted.) Regret is the psychological sensation of unhappiness from choosing X and foregoing a choice Y that turns out better. Regret theory predicts that a change in the statistical correlation of outcomes for two gambles X and Y can affect preferences, even if the probability distributions of X and Y are held constant. There is much evidence that correlation *does* matter, as regret theory predicts (Loomes, 1988; Loomes and Sugden, 1987b; Starmer and Sugden, 1987a; but see Battalio, Kagel, and Komain, 1988, table 7). In the experiment reported below, gamble outcomes are not statistically independent. The statistical correlation between gambles is held fixed so that the predictions of regret theory are the same as those of EU.[6]

1.3. Implicit expected utility

One way to generalize weighted utility is to allow the weighting function $W(X)$ used to evaluate a gamble to depend upon the weighted utility of the gamble. The same consequences result in expected utility if one allows the utility function used along an indifference curve to be different for each curve (i.e., to depend upon the expected utility on the curve).

Implicit weighted utility results if all gambles with weighted utility U^* will have the same $W(X, U^*)$ but those weights could vary with U^* (Chew, 1985). Then the utility of a gamble is given by U^*, which implicitly solves

$$\text{IWEU:}\quad U^* = U(qX + (1-q)Y) = \frac{qW(X,U^*)(U(X) + (1-q)W(Y,U^*)U(Y)}{qW(X,U^*) + (1-q)W(Y,U^*)} \tag{9}$$

This generalization is called *implicit* weighted utility, because U^* depends on the weighting function, which in turn depends on U^*. Implicit weighted utility results from a very weak version of the EU independence axiom: For each probability q and gamble Z, there is a probability r for which $X \sim Y$ implies $qX + (1-q)Z \sim rY + (1-r)Z$. Since r can depend on q and Z, it need not be independent of q and Z as in EU or independent of Z as in weighted utility.

In terms of the *meeting-point* property of weighted expected utility, implicit weighted utility has the odd property that the meeting point for indifference curves depends on the value of U^*; each indifference curve has its own meeting point (at which it may not meet any other curve), as in figure 3. Therefore, curves do not necessarily fan out or fan in uniformly.

The same result can be achieved by letting the utility function in EU depend on the expected utility U^* (i.e., replacing U(X) with $U(X, U^*)$, as in Dekel, 1986). Then EU is generalized to implicit EU,

$$\text{IEU:}\quad U^* = U(qX + (1-q)Y) = qU(X,U^*) + (1-q)U(Y,U^*). \tag{10}$$

Implicit EU makes the same prediction as implicit weighted EU: Indifference curves are straight lines (because Dekel (1986) assumes a *betweenness* axiom), but their slopes vary because $U(X, U^*)$ varies with U^*. Implicit EU describes a person who uses a different utility function, perhaps reflecting different degrees of risk aversion, along each indifference curve.

The only testable implication of implicit weighted utility and implicit EU (besides transitivity and continuity) is that indifference curves are straight lines. Linearity of curves arises from the betweenness axiom: If $X \prec Y$, then $X \prec qX + (1-q)Y\ Y$ for any probability q: i.e., any gamble that is a probabilistic mixture of X and Y should be between them in preference. Continuity and betweenness together mean that $X \sim Y$ implies $X \sim qX + (1-q)Y \sim Y$ for all q (i.e., people are indifferent to gambling amongst indifferent items). Since X and Y lie on the same indifference curve (by assumption), the mixture—which is indifferent to X and Y—must lie on that indifference curve too. But the probability mixture $qX + (1-q)Y$ is located on a line segment between X and Y; the indifference curve must be that line.

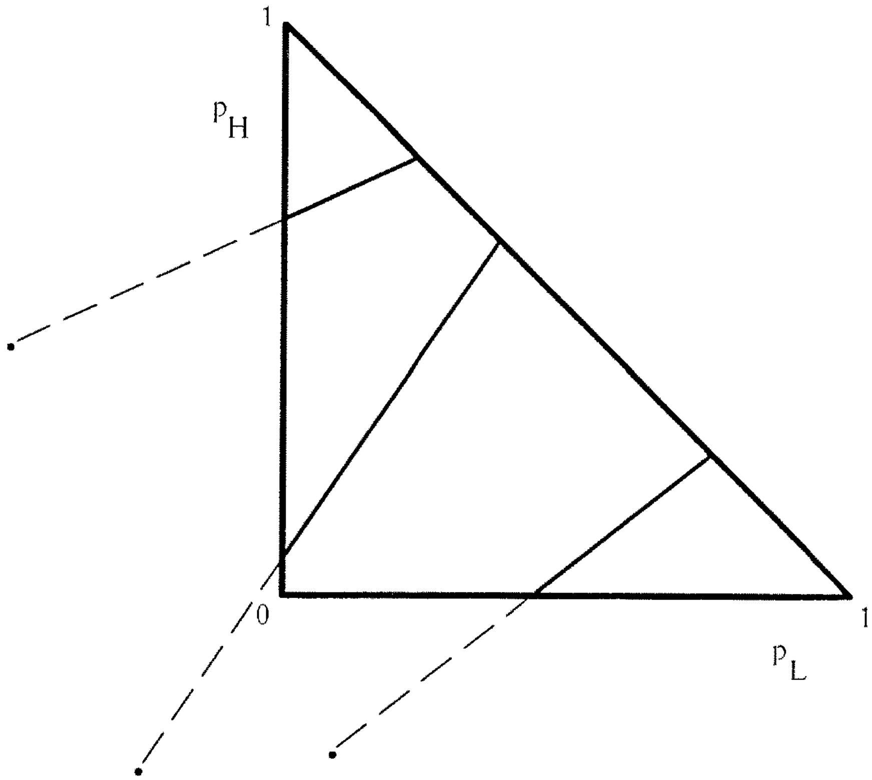

Fig. 3. Indifference curves assuming implicit weighted utility.

Since betweenness is the only testable property of implicit EU (and it is a testable implication of the independence axioms of EU and weighted utility also), it is fortunate that we can test it easily.

1.4. The fanning-out hypothesis

Machina (1982) argued that most empirical violations of the strong independence axiom of EU could be explained by the hypothesis that indifference curves fan out, or get steeper from the lower right to the upper left of the triangle. (Weighted utility, with $W(X_M) < 1$, also predicts fanning out, and implicit EU permits it.)

Machina's hypothesis uses the notion of (first-order) stochastic dominance. In general, a gamble X stochastically dominates Y if the cumulative distribution function (cdf) of X, denoted $F_X(w)$, is always below or the same as the cdf of Y (i.e., $F_X(w) \leqslant F_Y(w)$ for all w, with at least one strict inequality). For the three-outcome gambles shown in the triangle diagram, X stochastically dominates Y if $p_L(X) \leqslant p_L(Y)$ and

$p_H(X) \geqslant p_H(Y)$. Graphically, a point X stochastically dominates Y if X lies to the northwest of Y.

Machina considers preference functions $V(X)$ of a general form. He uses a first-order Taylor series expansion to show that preferences $V(X)$ for gambles near a point G are approximately the expectation of the *local* utility function at G, denoted $U(x;G)$: Expected utility holds locally even if it does not hold globally. (Similarly, expected value maximization holds locally in EU even if does not hold globally.) Machina then shows that many properties of the local utility functions of $V(X)$, such as risk aversion, imply global properties of $V(X)$; many of the standard assumptions of EU therefore hold for $V(X)$ even though $V(X)$ does not satisfy EU.

Machina's main theme is the theoretical demonstration that tools of EU analysis may be used despite EU violations, but he also made an important empirical conjecture. His claim, called *hypothesis II,* is that the local utility functions of stochastically dominant gambles will exhibit more risk aversion (by the Arrow-Pratt measure) than local utility functions of stochastically dominated gambles. Formally,

$$\text{If } F_G(w) \leqslant F_H(w)\ \forall\ w \text{ then } -U''(x;F_G)/U'(x;F_G) \geqslant -U''(x;F_H)/U'(x;F_H). \tag{11}$$

Since the steepness of indifference curves in the triangle diagram is a measure of risk aversion, hypothesis II predicts that curves will be steeper for gambles to the northwest (i.e., with lower p_L or higher p_H). Figure 4 shows curves that satisfy hypothesis II. Since such curves fan out, as in weighted utility (with $W(X_M) < 1$), we shall call Machina's hypothesis II the *fanning-out hypothesis.*

1.5. Lottery-dependent expected utility

Becker and Sarin (1987) describe a variant of EU in which the utility function used to evaluate outcomes depends on the particular lottery which generates those outcomes. Their *lottery-dependent EU* requires only ordering, continuity, and stochastic dominance axioms. Two further properties guarantee a functional form similar to implicit EU:

$$\text{LDEU:} \qquad \text{For } F = qX + (1-q)Y,\ U(F) = qU(X,c_F) + (1-q)U(Y,c_F), \tag{12}$$

where c_F is a number depending on F. (LDEU is easily generalized to gambles with several outcomes or continuous distributions.) If $c_F = U(F)$, LDEU is implicit EU; if c_F is constant, LDEU is EU. Becker and Sarin restrict their attention to a function c_F that is linear in probabilities, so $c_F = \int_x h(x)dF(x)$ for some function $h(x)$.

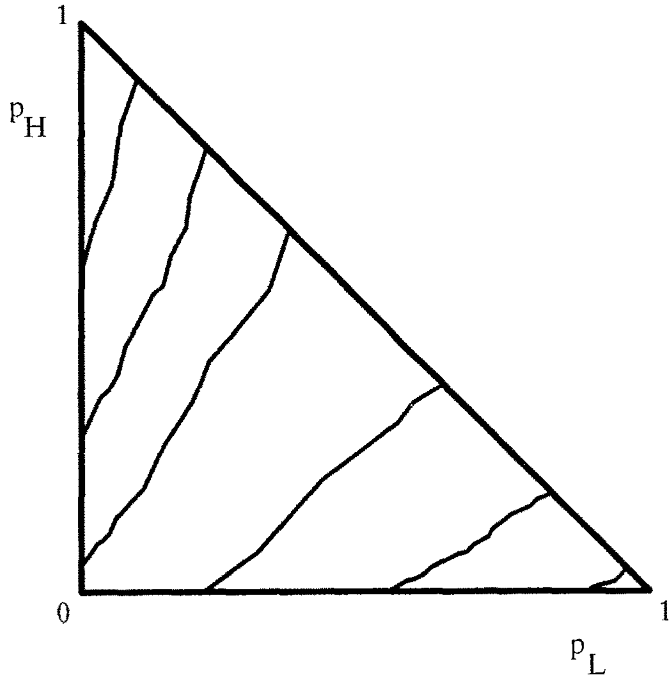

Fig. 4. Indifference curves assuming the fanning-out hypothesis.

(For instance, if $h(x) = x$, c_F is the mean of the distribution F.) They also suggest an exponential form for the utility function, $U(x) = (1 - e^{-c_F x'})/(1 - e^{-c_F})$ where $x' = (x - x_0)/(x^* - x_0)$, and x_0 and x^* represent the minimum and maximum values x can have.

Since lottery-dependent EU has no betweenness axiom, in its general form it has even fewer testable implications than implicit EU. However, the special cases with $c_F = \int_x h(x)dF(x)$ and exponential utility do make predictions about the shape of indifference curves. If we define $x_0 = X_L$ and $x^* = X_H$, the lottery-dependent utilities of our three-outcome gambles are

$$\text{LDEU:} \quad U^* = p_H + (1 - p_L - p_H)(1 - e^{-c_F X'_M})/(1 - e^{-c_F}) \tag{13}$$

For $X_M = (X_L + X_H)/2$ (as in the low gain and low loss gambles used in the experiment), (13) simplifies to

$$U^* = p_H + (1 - p_L - p_H)/(1 - e^{-c_F/2}) \tag{14}$$

$$\text{with } c_F = h(X_M) + p_L(h(X_L) - h(X_M)) + p_H(h(X_H) - h(X_M))$$

Since (14) is not a linear function of the probabilities, solving for p_H in terms of p_L

to see the shape of indifference curves gives complicated expressions that are hard to interpret. However, we can take the derivative dU^*/dp_L from (14), note that $dU^*/dp_L = 0$ along an indifference curve, and solve for the slope of indifference curves dp_H/dp_L. We get

$$\frac{dp_H}{dp_L} = \frac{1 + e^{-c_F/2}[1 + .5(h(X_M) - h(X_L))(1 - p_L - p_H)]}{e^{-c_F} + e^{-c_F/2}[1 + .5(h(X_H) - h(X_M))(1 - p_L - p_H)]}. \quad (15)$$

From (15) we can intuit some properties of indifference curves:[7] If $h(x)$ is linear in x, then for the special case $X_M = (X_L + X_H)/2$ it is clear that $h(X_M) - h(X_L) = h(X_H) - h(X_M)$. As p_L increases with p_H held fixed, c_F decreases, e^{-c_F} increases, and dp_H/dp_L decreases. By similar reasoning, increases in p_H cause dp_H/dp_L to increase. Therefore, indifference curves get steeper as p_H increases along the left edge $p_L = 0$, and flatter as p_L increases along the lower edge $p_H = 0$: the curves fan out.

Becker and Sarin show that if $h(x)$ is positive and concave then people are risk-averse; convexity implies risk-seeking. If $X_M = (X_L + X_H)/2$ and p_L increases, concavity implies that $h(X_M) - h(X_L)$ is greater than $h(X_H) - h(X_M)$. For differential increases in both p_H and p_L, it follows from (15) that concavity implies decreases in dp_H/dp_L as we move up an indifference curve (convexity implies increases).[8] That is, the curves will be concave if $h(x)$ is concave, and convex if $h(x)$ is convex.

Figure 5 shows typical indifference curves for lottery-dependent EU, assuming the exponential form with $c_F = \int_x h(x)dF(X)$, $h(x)$ positive and concave. They fan out, as curves do in weighted utility (or by the fanning-out hypothesis), but they are also concave (or convex, depending on $h(x)$) like curves in the rank-dependent utility theories discussed below.

1.6. *Prospect theory*

The prospect theory of Kahneman and Tversky (1979) does not generalize EU. It embodies four important differences from EU: (1) Prospect theory only applies to *prospects,* gambles with at most two nonzero outcomes (though see Tversky and Kahneman, 1987, footnote 3); (2) prospects are edited to make them simpler to evaluate (e.g., outcomes and probabilities are rounded off or lumped together); (3) all outcomes are framed as changes from a reference point; and (4) edited prsopects are evaluated according to one of several expectationlike rules that combine a value $v(X)$ and a decision weight $\pi(q)$. For instance, if X and Y are outcomes of the same sign (both gains, or both losses, relative to the reference point), then

$$U(r0 + qX + (1 - r - q)Y) = \pi(q)v(X) + \pi(1 - r - q)v(Y) \quad (16)$$

for $r > 0$. The decision weight function $\pi(q)$ transforms probabilities nonlinearly. Kahneman and Tversky report data which suggest that $\pi(q)$ is increasing, subadditive ($\pi(q) + \pi(1 - q) < 1$), and discontinuous at the endpoints 0 and 1.[9] They also

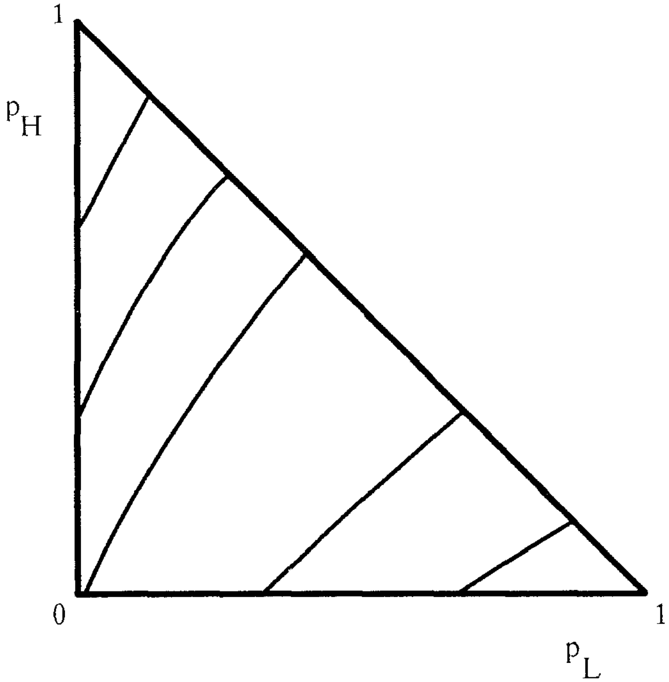

Fig. 5. Indifference curves assuming lottery-dependent expected utility ($h(x)$ concave).

hypothesize that the value function $v(x)$ is concave for gains (reflecting risk aversion), convex for losses (reflecting risk-seeking), and steeper for losses than for gains (reflecting loss aversion).

Prospect theory is difficult to test because it has many more degrees of freedom, especially in the editing stage, than any other theory. The whole idea of an indifference curve is jeopardized by the possibility of editing and framing. However, we can hope to learn something by using suitably limited prospects (satisfying (i) above), leaving parts (ii) and (iii) aside, and testing whether the evaluation rules and hypothesized $v(x)$ and $\pi(q)$ functions describe choices. The evaluation rules may work well even if editing and framing are ignored. If the rules work poorly, the nature of their descriptive errors may suggest the kind of editing and framing subjects are doing.

In the tests we used only gambles with a zero outcome and (at most) two nonzero outcomes, so prospect theory applies. The gambles were of the form $0 = X_L < X_M < X_H$, or $X_H < X_M < X_L = 0$. When $p_L = 0$, along the left edge of the triangle, prospects are assumed to be edited into a sure-gain component X_M and a risky component $p_M 0 + p_H(X_H - X_M)$ (as suggested by the theory). These edited prospects have value

$$\text{PT}, p_L = 0: \qquad U(p_M X_M + p_H X_H) = v(X_M) + \pi(p_H) v(X_H - X_M). \tag{17}$$

When $p_L > 0$, prospects have value

$$\text{PT}, p_L > 0: \qquad U(p_L 0 + p_M X_M + p_H X_H) = \pi(p_M)v(X_M) + \pi(p_H)v(X_H). \tag{18}$$

As we did for lottery-dependent utility above, we can solve for dp_H/dp_L by differentiating (15) with respect to p_L. We get

$$\frac{dU(p_L 0 + p_M X_M + p_H X_H)}{dp_L} = \frac{d\pi(1 - p_L - p_H)}{dp_L} v(X_M) + \frac{d\pi(p_H)}{dp_L} v(X_H). \tag{19}$$

Along an indifference curve, $U(p_L 0 + p_M X_M + p_H X_H)$ is constant so its derivative is zero. Applying the chain rule to the other terms, we get

$$dp_H/dp_L = \pi'(p_M)v(X_M)/(\pi'(p_H)v(X_H) - \pi'(p_M)v(X_M)). \tag{20}$$

Equation (20) tells us the slope of the indifference curve at any point, in terms of the derivatives of the decision weight function $\pi(q)$ and the values $v(X_M)$, $v(X_H)$. (We must be careful in applying (20) because derivatives of $\pi(q)$ do not exist on the edges of a triangle where one of the three probabilities is zero, if $\pi(\cdot)$ is not differentiable at zero.)

The properties of the $\pi(q)$ function posited by Kahneman and Tversky suggest, but do not necessarily imply, that $\pi(q)$ is convex (except near the endpoints). We take convexity to be an adequate working assumption. If $\pi(q)$ is convex, its derivative is smallest around zero and largest around one. The slope dp_H/dp_L is maximized—indifference curves will be steepest—near the lower left-hand corner where p_M is close to 1. As p_H increases near the left edge, the slope dp_H/dp_L gets smaller (indifference curves will be flatter); as p_L increases on the lower edge the slope gets smaller, but at a slower rate than on the left edge. Near the hypotenuse, with p_M near zero and fixed, the slope gets flatter as p_H increases. Notice that indifference curves fan out through part of the triangle, near the edge $p_H = 0$ for instance, but they fan in through other parts, near $p_L = 0$ or $p_M = 0$.

Because $\pi(q)$ is discontinuous near 0 and 1, the indifference curves are discontinuous along the edges, as figure 6 shows. Because people will overweight the small probabilities of winning X_M (near the hypotenuse) or X_H (near the lower edge), they will much prefer points just inside those edges to nearby points that are exactly on the edge: indifference curves on the hypotenuse and lower edge will appear very steep and flat, respectively. On the left edge there is a change in evaluation rule, from (16) to (17), as p_L goes from zero to a small number. This change makes the edge points where $p_L = 0$ much preferred to points inside the edge, so indifference curves will appear very steep on the left edge of the triangle.

Along the line $p_H = p_M$ (or $p_H = (1 - p_L)/2$), which bisects the triangle, the slope dp_H/dp_L is constant; indifference curves are exactly parallel along this line. We call this property *bisector parallelism*. These properties are reflected in the indifference

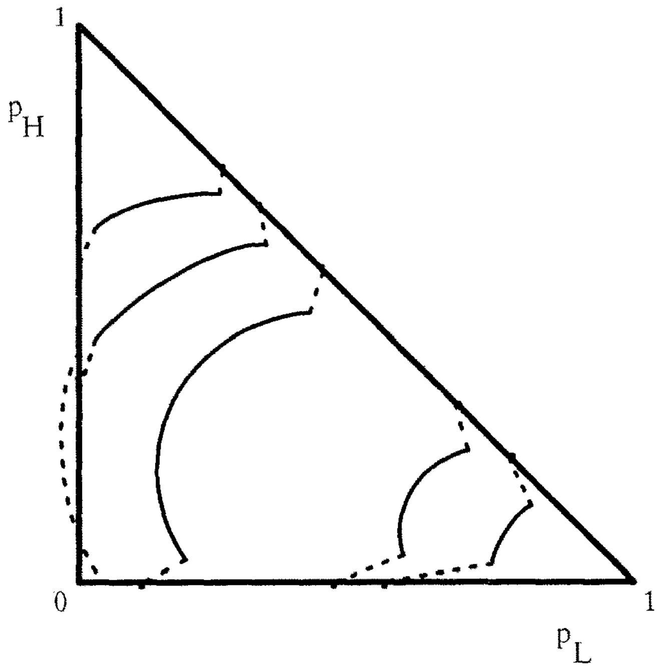

Fig. 6. Indifference curves assuming prospect theory.

curves graphed in figure 6.[10] All these properties depend only on the convexity of the decision weight function.

There are many other theories in which probabilities are weighted nonlinearly, or subjectively (see Edwards, 1954a; Handa, 1977; Karmarkar, 1978). Their indifference curves will still have slopes given by (20). Specific predictions depend upon the shape of the subjective weighting functions. For instance, if people attach special weight to some probability p (Edwards, 1954b) then indifference curves will be kinked along the horizontal line $p_H = p$ and the diagonal line $p_M = p$. (Our test is probably not subtle enough to detect such kinks.) If subjective probabilities of gains are overweighted compared to identical probabilities of losses (Edwards, 1955), the degree of curvature in indifference curves will differ for gains and losses.

1.7. Expected utility with rank-dependent probabilities

In prospect theory, nonlinear weighting of probabilities can cause violations of stochastic dominance. (We can see this in the negatively sloped segment of indifference curve in the lower left-hand corner of figure 6. Along this segment, some gambles stochastically dominate others but all are equally preferred.) In prospect theory, people are assumed to notice dominance in the editing stage and choose the dominant prospect, but this presumption can cause indirect transitivity violations. Tversky and Kahneman (1987) argue that dominance violations do occur, and a theory which tries to account for them is better than a theory which rules them out.

Several authors have axiomatized generalized utility theories in which probabilities are weighted nonlinearly but dominance is not violated. The trick is to take the cumulative distribution function (cdf) of each gamble and transform the entire cdf by a nondecreasing function $g(q)$. If X stochastically dominates Y, then $F_X(x) \leqslant F_Y(x)$ for all x and $g(F_X(x)) \leqslant g(F_Y(x))$ too. Outcomes are then weighted by discrete chunks of the cdf (or differentials if the cdf is continuous).

These theories weight probabilities of outcomes according to the order or rank of the outcomes in the transformed cdf. Therefore, they are sometimes called EU theories with *rank-dependent probabilities* (EURDP). In general, expected utility with rank-dependent probabilities is

$$U(qX + (1 - q)Y) = g(q)U(X) + (1 - g(q))U(Y) \qquad \text{for } X < Y. \tag{21}$$

We assume $g(q)$ is monotonically increasing, with $g(0) = 0$ and $g(1) = 1$. If $g(q) = q$ then EURDP reduces to EU.

Quiggin (1982,1985) was the first to consider such a theory (called *anticipated utility*), in which $g(.5) = .5$. Yaari (1987) considers a special case in which $U(x)$ is a linear function of x. He locates risk aversion in the probability weights (assuming $g(q)$ is concave[11]) rather than in the utility function, so he calls his theory *dual EU*. Hey (1984) interpreted the probability weights that result from a concave $g(q)$ as reflections of optimism and pessimism. Segal (forthcoming) used the theory to explain aversion to ambiguity about probabilities. Chew (1984) axiomatized the general form given in (21) above.

In our three-gamble example, EURDP states that

$$U^* = g(p_L)U(X_L) + [g(p_L + p_M) - g(p_L)]U(X_M) + (1 - g(p_L + p_M))U(X_H). \tag{22}$$

As with prospect theory, we can derive the slope of indifference curves dp_H/dp_L by differentiating with respect to p_L, noting that $dU^*/dp_L = 0$ along an indifference curve, and applying the chain rule. We find (see also Roell, 1987, p. 156)

$$dp_H/dp_L = \frac{g'(p_L)(U(X_M) - U(X_L))}{g'(1 - p_H)(U(X_H) - U(X_M))}. \tag{23}$$

The properties of dp_H/dp_L in EURDP are quite similar to those in prospect theory. Typical indifference curves (for a concave $g(q)$) are shown in figure 7. If $g(q)$ is concave (convex), indifference curves are steepest (flattest) in the corner $p_M = 1$; they

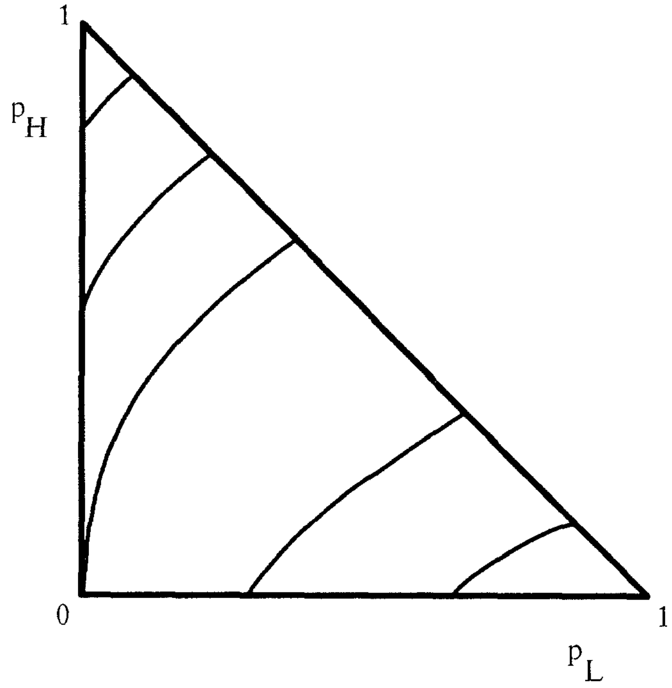

Fig. 7. Indifference curves assuming expected utility with rank-dependent probabilities ($g(q)$ concave).

get flatter (steeper) as one moves along the left edge $p_L = 0$ or the lower edge $p_H = 0$. The curves are equal in slope along the hypotenuse $p_M = 0$ (where $p_L = 1 - p_H$) for *any* $g(q)$. We call the latter property *hypotenuse parallelism*. As in prospect theory, the curves do not fan out uniformly from lower right to upper left.

2. Experimental design

Each subject was shown several pairs of gambles from a set of 14 pairs, which are given in table 2 and shown graphically in figure 8. Each pair was a choice between a gamble $G = (p_L,p_M,p_H)$ and a transformation of G with some of the probability mass p_M shifted from the middle outcome X_M to each of the extreme outcomes X_L and X_H. Either .10 or .20 probability mass was shifted, so the choices were either between G and $(pL + .10,p_M - .20,p_H + .10)$ or G and $(pL + .20,p_M - .40,p_H + .20)$. The gambles shifted .10 are numbered above the lines in figure 8; gambles shifted .20 are numbered below the line.

Three payoff levels (X_L,X_M,X_H) were used: large gains (0;$10,000;$25,000); small gains (0;$5;$10); and losses (−$10;−$5;0). These payoffs enable us to test whether

Table 2. Gamble pairs presented to subjects

Pair no.	Less risky gamble P_L	P_M	P_H	More risky gamble P_L	P_M	P_H
1	0	.2	.8	.1	0	.9
2	0	.6	.4	.1	.4	.5
3	0	.6	.4	.2	.2	.6
4	.1	.4	.5	.3	0	.7
5	0	1.0	0	.1	.8	.1
6	0	1.0	0	.2	.6	.2
7	.3	.4	.3	.5	0	.5
8	.4	.2	.4	.5	0	.5
9	.4	.6	0	.5	.4	.1
10	.4	.6	0	.6	.2	.2
11	.5	.4	.1	.7	0	.3
12	.8	.2	0	.9	0	.1
13	.2	.2	.6	.3	0	.7
14	.6	.2	.2	.7	0	.3

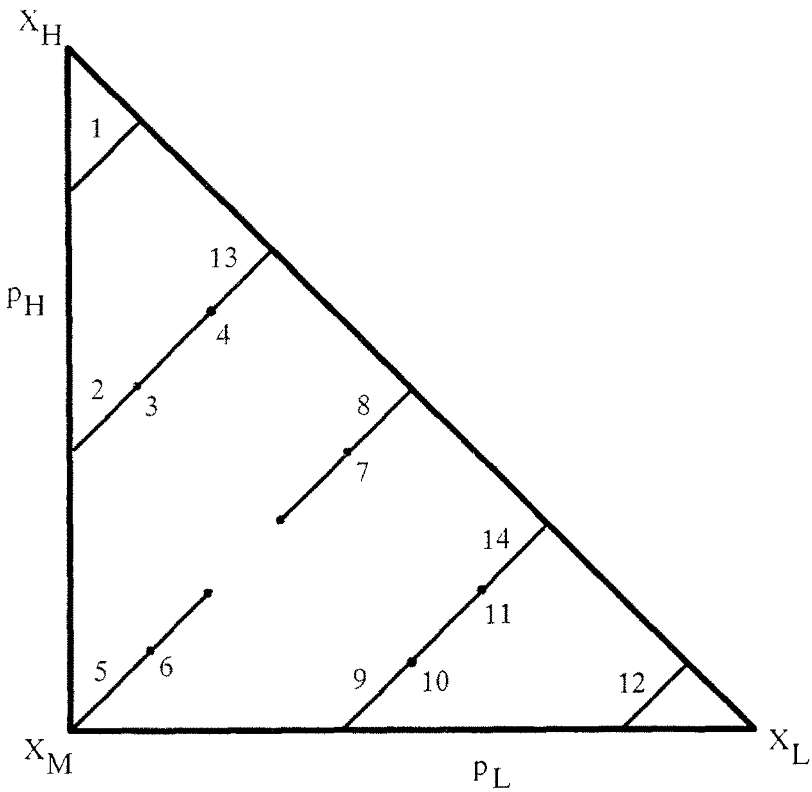

Fig. 8. Gamble pairs presented to subjects.

violations of EU vary with the size of the payoffs (as weighted utility, fanning out, and lottery-dependent utility allow), and whether choices for losses are fundamentally different than choices for gains (as prospect theory suggests).

By construction, the gambles in each small gain and loss pair always had the same expected value. Therefore, the shift in probability mass away from p_M (which characterized the change between gamble G and gamble H in each pair) is a mean-preserving spread. For large gains, gambles G and H do not have the same expected value, so the shift away from p_M is not a *mean-preserving* spread, but it does increase variance (i.e., the variance of H is larger than the variance of G). Therefore, we will call the original gamble G in each pair *less risky* and the transformed gamble H *more risky*.

Every subject made choices for four pairs of gambles in each of the three payoff levels, and a single repetition of one of the 12 pairs, a total of 13 choices. For a given payoff level, the four gambles were chosen to test betweenness and fanning out for each subject.

Betweenness was tested by giving subjects pairs of gambles of the sort $G = (p_L,p_M,p_H)$ vs. $I = (p_L + .20, p_M - .40, p_H + .20)$ and G vs. $H = (p_L + .10, p_M - .20, p_H + .10)$. Since $H = .5G + .5I$, if subjects obey betweenness the only permissible preference pairs are $(H \prec G, I \prec G)$ and $(G \prec H, G \prec I)$. For example, consider pair 2, the choice between (0,.4,.6) and (.1,.2,.7), and pair 3, the choice between (0,.4,.6) and (.2,0,.8). A coin flip between (0,.4.,.6) and (.2,0,.8) gives (.1,.2,.7); therefore if people prefer (0,.4,.6) to (.2,0,.8) and obey betweenness, then they should prefer (0,.4,.6) to (.1,.2,.7).

Fanning out was tested by giving subjects gamble pairs in different parts of the triangle. For instance, a subject might choose among gambles in pair 2, 5, 6, and 9. Since the gambles in pair 2 dominate the gambles in pair 5, which dominate those in pair 9, fanning out predicts that preferences for the riskier gamble in each pair will increase from pair 2 to pair 5 to pair 9. (Pairs 5 and 6 were used only to test betweenness, though pair 6 also provides data about fanning out.)

Each pair of gambles was represented on a separate sheet of paper using a diagram shown (for gamble pair 9) in the appendix. Gambles were operationalized as random drawings of tickets that were uniformly distributed from 1 to 100. The numbers of tickets yielding a particular prize were shown on the vertical axis. Amounts of prizes were shown in rectangles, with the width of the rectangles drawn to reflect the amount of the prize. In these diagrams the area of the rectangles is the (relative) expected value of a gamble, so subjects can judge expected value of gambles visually. The diagrams might bias the results in favor of expected value maximization and hence in favor of EU rather than competing theories.[12] The bias is deliberate: it helps rule out the criticism that subjects only violate EU because they are not told expected value or cannot calculate it.

Each subject received a questionnaire with two pages of instructions (see appendix), 13 pages with one gamble pair on each page, and a page for recording their payoff. We randomly varied the order of gamble pairs, and whether the less risky gamble appeared on the left or right. (Subjects chose gambles on the right side of the page 50.77% of the time, $n = 4103$, $\chi^2 = .96$.)

The experiment was run in seven sections of an MBA quantitative methods course at Wharton.[13] One hundred and twenty of the original subjects were later asked four questions each (gamble pairs 13–14, and some replications of 3–4), without payment, about a month after the initial experiment.

There were three payment conditions: (1) 179 subjects did not play any of the gambles, but were paid $2 for participating; (2) 80 subjects played one of the small gain gambles; and (3) 96 subjects were given $10, then played one of the loss gambles (with a maximum loss of $10). Comparing choices in condition (1) with choices in conditions (2–3) tests whether paying the subjects motivates them to choose differently (perhaps more carefully, or more in accordance with EU).

After subjects made all their choices, they played gambles by choosing a ticket to determine which gamble pair was chosen, and a ticket to determine the outcome of the gamble they preferred in the chosen pair.[14] After picking the ticket that determined the gamble pair, half of the subjects who actually played gambles (conditions 2–3) had a chance to change their choices. After subjects decided whether to change or not, they chose the second ticket, which determined their payoffs.[15]

First we will draw methodological conclusions, and then the theories will be tested.

3. Results: Methodological conclusions

3.1. Reliability

We can test the reliability of subjects' responses by seeing how often they expressed the same preference for the same gamble. Since subjects were not allowed to express indifference,[16] unreliability consists of human error in expressing true preferences, and random switches between A and B when the true preferences is indifference.

Overall, 31.6% ($n = 348$) of the subjects reversed preference. This number is distressingly close to the 50% we would expect if choices were random, but comparable with numbers in other studies (e.g., 26.5% in Starmer and Sugden, 1987b).

The fraction of reversals did not depend on whether subjects actually played gambles ($\chi^2 = .22$) or whether gambles were presented on different sides of the page in the two replications ($\chi^2 = .43$). However, the fraction of reversals was lower when there were fewer gamble pairs between the first and second presentations of the repeated gamble (suggesting that poor memory contributes to unreliability).[17] For instance, when pairs were presented next to each other, there were only 13.3% reversals ($n = 30$); when pairs had 0–3 other pairs between them, there were 24.1% reversals ($n = 112$).

For the tests of EU involving gamble pairs that came next to each other in the sequence of pairs, we can take 13.3% reversals as a baseline of allowable error. For pairs less than four pairs apart, the more conservative baseline of 24.1% is appropriate.

3.2. Reluctance to change decisions: The isolation effect

Eighty subjects were allowed to change their expressed preference after the gamble pair they would actually play was determined. Allowing subjects to change their choices tested whether subjects choose between gambles G and H as if the G–H pair is certain to be chosen, or between $\frac{1}{4}G$ and $\frac{1}{4}H$ (since there is only a one-in-four chance that the G–H pair will actually be chosen to be played). If subjects do change, their preferences for a choice between $\frac{1}{4}G$ and $\frac{1}{4}H$ are apparently different than their preferences between G and H. Subjects will not change if their preferences obey the independence axiom, or if they are *dynamically consistent* and choose as if each gamble will actually be selected (as in the *isolation effect* of prospect theory; see Kahneman and Tversky, 1979, and Loomes and Sugden, 1986). A change in preferences is evidence of an EU-violating *common ratio effect* (see below).

Only two of 80 subjects did change.[18] Therefore, either the independence axiom holds or subjects exhibit an isolation effect. Since the data below suggest that independence is often violated, we must conclude that there is an isolation effect. This is puzzling for theorists, but comforting for experimenters because it implies that allowing subjects to play some randomly chosen gambles generates meaningful responses for *all* gambles.

3.3. The effect of incentives

Financial incentives are controversial because social scientists often disagree about their effect. Noneconomists usually assume that financial incentives do not matter much because subjects have no reason to misrepresent their preferences and are motivated by other incentives. Economists believe instinctively that financial incentives matter because they relate outcomes to decisions most saliently (cf. Smith, 1982); subjects who are not financially motivated may be more inclined to give answers that are quick or sloppy or amusing (Grether, 1981), or they may give answers they think experimenters want to hear. But it is unclear why such answers would necessarily violate EU (provided the number of violations are compared to subject reliability).

Subjects who actually played a gamble were no more reliable than subjects who did not play, and they took the same amount of time making choices. It is also useful to see whether financial incentives affected the percentage of subjects choosing

Table 3. Chi-squared tests of independence of choices from payment condition

Gamble pair	Type of payoff		
	Large gain	Small gain	Loss
1	2.68	3.90	.48
2	1.70	1.13	4.57
3	3.20	.83	3.90
4	2.53	1.34	6.35*
5	3.40	1.44	.16
6	1.69	1.60	2.58
7	1.23	2.69	10.11**
8	.28	2.52	4.04
9	.01	.81	1.29
10	1.14	.42	5.87
11	.74	2.41	6.04*
12	3.57	11.20**	2.09
Summed statistics	22.17	30.29	82.58**

*$p < .05$.
**$p < .01$.

the less risky gamble (the lower-left gamble in each pair) in each of the three incentive conditions.

For each of the 12 gambles administered in the initial, large experiment, a contingency table can be constructed showing the numbers of less risky and more risky choices (rows) in each of the three payment conditions (columns)—no gamble, gain gamble, loss gamble. A chi-squared statistic (with two degrees of freedom) can be used to test for independence of the fraction of less risky choices from the incentive conditions. These statistics are shown in table 3, for each of the 12 gambles, for each of the three levels of payoffs. Summed chi-squared statistics are shown at the bottom of each column; these have the chi-squared distribution with 24 degrees of freedom if incentives do not matter.

Most of the chi-squared statistics in table 3 are well within the bounds for acceptance of the null hypothesis of independence. For the large and small gain payoffs, there is only one strong rejection (at the 1% level) out of 24. Incentives make little difference in subjects' choices among gambles involving gains.

3.4. Gambles for losses from stakes vs. gambles for gains

For loss payoffs, there are three rejections in table 3, and the summed statistic (82.58) strongly rejects independence. These rejections are problematic because subjects who actually played loss gambles were given \$10 initially; they might consider their gambles to be small *gain* gambles. To test this possibility, we can compare the choices of subjects who played loss gambles with choices on *gain* gambles that were played. We can also compare choices on played loss gambles with choices on unplayed loss gambles. These two comparisons are shown in table 4. If subjects playing loss gambles treat them like losses, the chi-squared statistics in the left column will be large and those in the right column will be small. If they treat played loss gambles like net gains (which they are), the left-column statistics will be small and the right-column statistics will be large.

All the chi-squared statistics are rather large; choices involving losses that will be played are apparently different from both unplayed loss choices and played gain choices.

Battalio, Kagel, and Komain (1988) also found that losses from a stake were treated differently than equivalent gains (their table 3). However, their subjects also made more risk-averse choices when they actually played gambles than when they made hypothetical choices (though the direction of majority risk preference was rarely changed by playing).

The mixed evidence in table 4 suggests that some subjects may treat losses from a stake like gains and some may treat them like losses. We can roughly measure how many subjects think either way by looking at each individual subject's pattern of choices. Table 5 summarizes the number of less risky choices in small gain gamble pairs and the number of less risky choices in loss gamble pairs for each

Table 4. Chi-squared tests of played loss choices versus hypothetical gain and loss choices

	Played loss choices vs.	
Gamble pair	Hypothetical small gain	Hypothetical loss
1	.09	.44
2	.12	4.57*
3	7.33**	3.56
4	1.34	6.35*
5	1.94	.07
6	6.25*	.19
7	2.46	9.94**
8	1.09	1.66
9	.62	.07
10	9.43**	5.74*
11	12.33**	5.46*
12	.24	1.44
Summed statistics	43.24**	39.49**

*$p < .05$.
**$p < .01$.

Table 5. The number of less risky choices on gain and loss gambles by each subject

		Number of less risky choices on LOSSES					
		0	1	2	3	4	Total
Number of less risky choices on GAINS	0	5	0	0	3	0	8
	1	0	4	4	1	2	11
	2	4	9	3	3	2	21
	3	5	9	10	3	1	28
	4	8	8	4	4	4	28
	Total	22	30	21	14	9	96

χ^2 statistic for collapsed table (1–3 combined) = 11.29 ($p < .025$).

subject who played a loss gamble. Nineteen subjects (those on the diagonal) made the same number of less risky choices for gains as for losses. Sixteen subjects were above the diagonal, more risk-averse for losses than for gains, and 61 subjects were below the diagonal, more risk-averse for gains than for losses.

Table 5 shows that a large majority of subjects treat losses from a stake as different from gains, and are more risk-averse toward gains than toward losses. The *reflection effect* in prospect theory suggests a stronger claim, that subjects will be risk-averse toward gains and risk-seeking toward losses. Thirty of 57[19] subjects (53%) exhibited reflection by making a majority of risk-averse gains choices and a majority of risk-preferring loss choices. This percentage is roughly comparable to the figures in Hershey and Schoemaker (1980), Chew and Waller (1986), and Battalio, Kagel, and Komain (1988, table 2; cf. table 1). Risk aversion toward gains and risk preference for losses does not hold for all subjects, but it is the modal preference pattern among the four possible patterns.

4. Results: Tests of competing theories

There are two ways to analyze the data: between-subjects and within-subjects. Between-subjects tests look at patterns of averaged choices; within-subjects tests look at averaged patterns of choices. Within-subject analyses are generally preferable because they do not require the assumption that subjects' tastes (including tendencies to violate EU) are the same except for random deviations. Between-subjects analyses do require that assumption. We shall use between-subjects analyses to suggest conclusions that will be verified by within-subjects analyses. Between-subjects analyses are also useful because many applications of EU in economics assume a *representative agent;* between-subject measurements of average behavior provide a picture of how such a hypothetical representative agent might act.

4.1. Between-subjects analyses

The fractions of all subjects[20] choosing the less risky gamble in each pair are shown graphically in figures 9–11. Each figure is a triangle diagram with the amounts X_L, X_M, X_H shown on the corners. The thin lines connect the two gambles in a pair. (For the small gain and loss gambles, figures 10–11, these are isoexpected value lines.) The thick line represents the fraction of subjects who chose the less risky gamble in the pair (the fraction is written next to the thick line). For instance, the thin line in the upper left-hand corner of figure 9 connects (0,.2,.8) and (.1,0,.9),

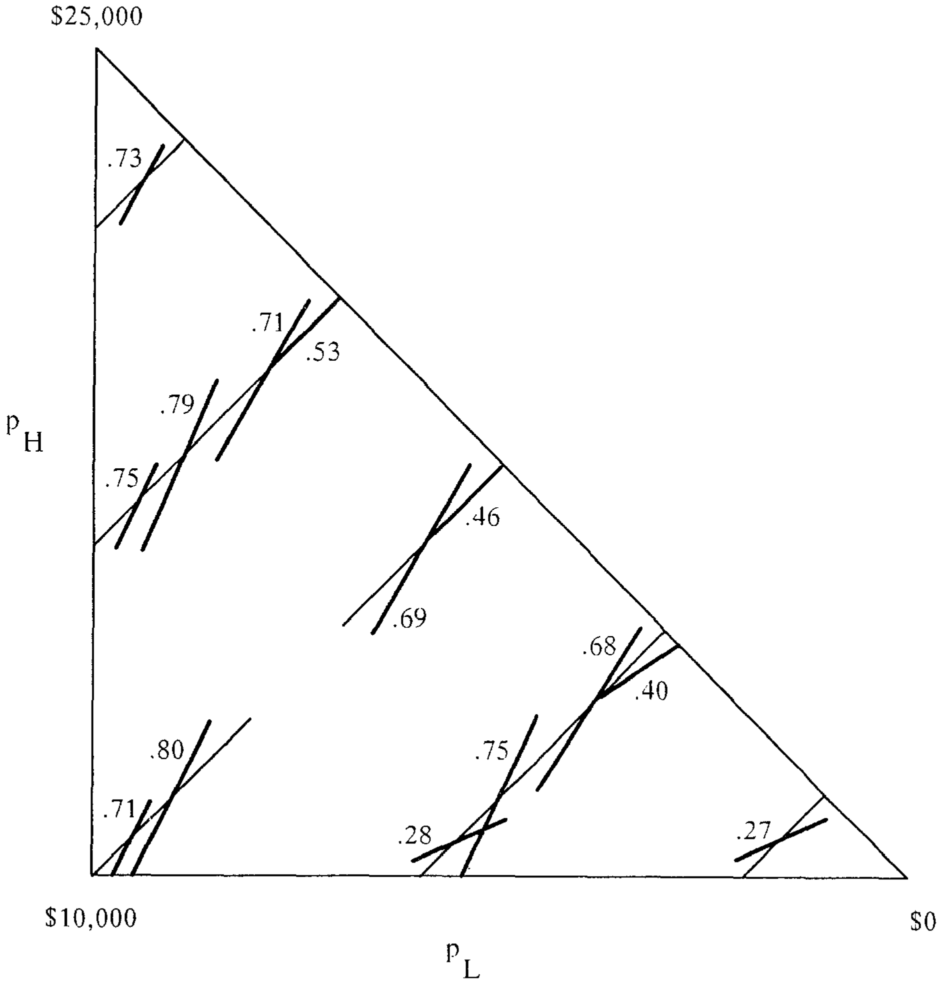

Fig. 9. Fraction of subjects choosing the less risky gamble in each pair, large gains.

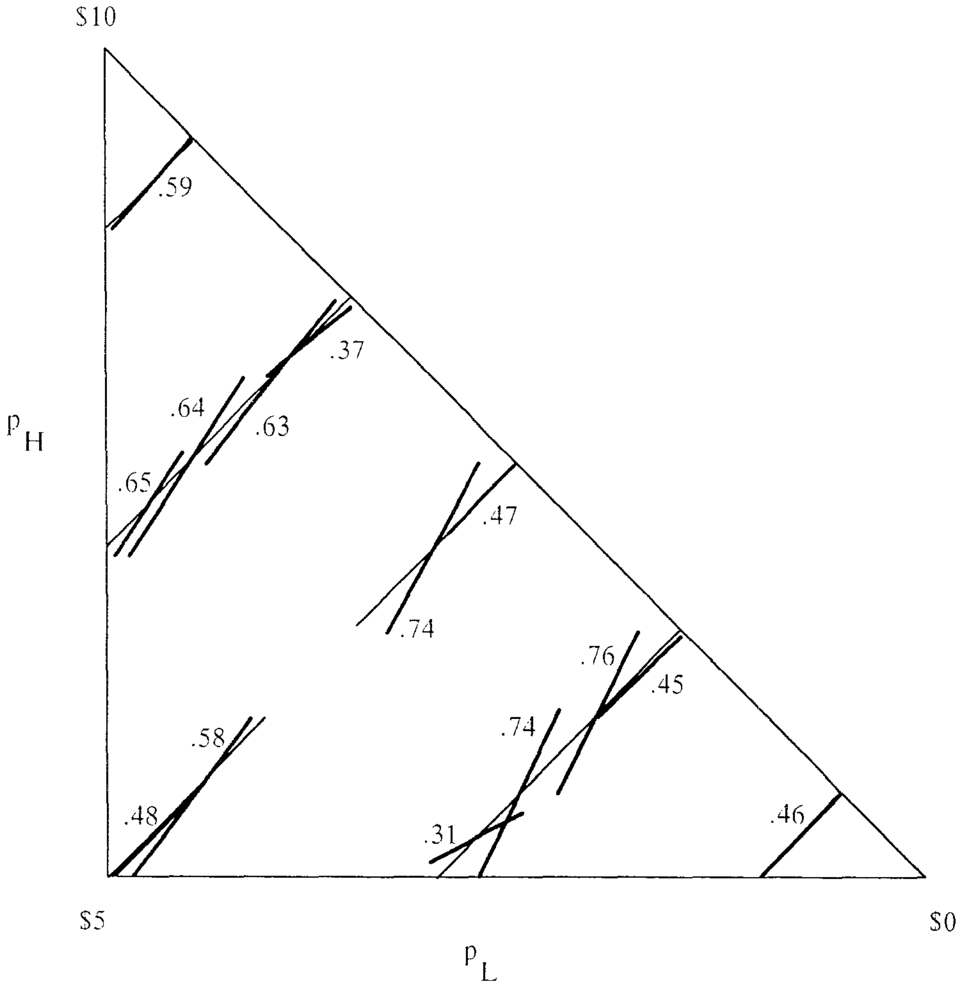

Fig. 10. Fraction of subjects choosing the less risky gamble in each pair, small gains.

the two choices in gamble pair 1. Seventy-three percent of the subjects chose (0,.2,.8) over (.1,0,.9).

The slope of the thick line is a linear function of the fraction of subjects who chose the less risky gamble. If all subjects chose the less risky gamble, the thick line will be perfectly vertical; if all chose the more risky gamble, it will be horizontal. If half chose the more risky gamble and half chose the less risky gamble, the thick line will have a slope of one (it will lie on top of the thin line connecting the gambles). The thick lines are analogous to indifference curves, but they have no formal meaning.[21] Iron filings scattered on paper will line up when a magnet is held underneath the paper, revealing the direction of the magnetic field. The lines

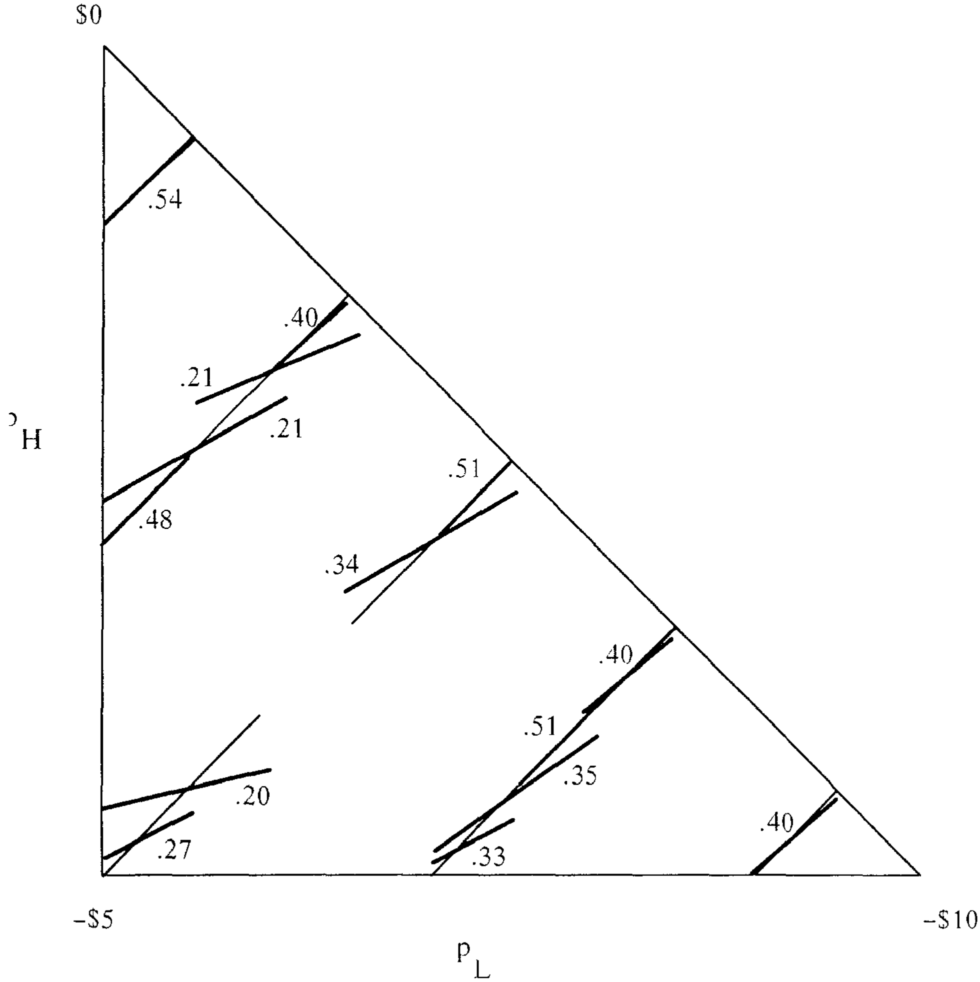

Fig. 11. Fraction of subjects choosing the less risky gamble in each pair, losses.

in figures 9–11 are our iron filings; they suggest the direction of EU violations, which can then be tested with within-subjects data (which *do* have formal meaning).

EU predicts the lines will be parallel. That is a reasonable approximation, especially for the gambles involving small money amounts (figures 10–11). However, a difference between two fractions of less risky choices of roughly .10 is statistically significant at the 5% level,[22] and there are many such differences. We can strongly reject the hypothesis that the fractions in each triangle are all equal.[23]

The important question is whether any of the generalized utility theories can account for the deviations from EU. We can get a rough idea of how well the generalized theories fare by comparing the properties they predict curves should have, shown in table 1, with the thick lines in figures 9–11.

Because weighted and implicit EU use the betweenness axiom, they predict that different gamble pairs on a single thin line will have thick lines with the same slope. This property often holds, but it is strongly violated for gamble pairs 9–11 and 14 in the lower right portion of figures 9–10 (gains), and for gamble pairs 2–4 and 13 in figure 11 (losses).

Fanning out holds reasonably well: indifference curves typically get steeper as one goes from lower right to upper left (less so in figure 11, for losses). Gamble pairs near the hypotenuse do seem to fan out less than along other edges, as predicted by EURDP.

Curves are different for small gains (figure 10) and small losses (figure 11). The loss gambles are identical to the gain gambles except for a difference of $10. If utility functions were defined on final wealth positions, the difference of $10 would be a minor wealth effect and figure 10 would look like figure 11. It does not. Subjects seem to value gains and losses from a reference point, as many theorists (e.g., Markowitz, 1952; Kahneman and Tversky, 1979) have suggested.

The steepness of the gain curves (figure 10) and the flatness of the loss curves (figure 11) suggest that subjects are risk-averse for gains and risk-seeking for losses, as in prospect theory (see also Battalio, Kagel, and Komain, 1988; cf. Harless, 1988). Indifference curves also seem to be mostly convex for small losses and concave for small gains, which is inconsistent with EURDP but consistent with prospect theory. Indeed, the curves in the loss gambles (figure 11) are much like the curves from the small gain gambles (figure 10) reflected around the 45-degree line, as prospect theory predicts.

Prospect theory can account for discontinuities at the edges. Discontinuities imply that gamble pairs like 2 and 3, which include the same edge gamble and two linear gambles which have p_H and p_L larger by .1 (pair 2) and .2 (pair 3), will have quite different slopes (recall figure 6). Such pairs usually do have different slopes. However, prospect theory predicts that the pairs closer to the hypotenuse (e.g., pair 8 and pair 14) should have *steeper* slopes than their associated inner pairs (pairs 7 and 11), but the closer pairs actually have flatter slopes. Except for these hypotenuse pairs,[24] prospect theory explains the basic features of figures 9–11 rather well.

The impressions one gets from the between-subjects analyses also hold within subjects.

4.2. *Within-subjects analyses*

We can test betweenness within-subjects by looking at the fraction of subjects who choose the less or more risky gamble both times from pairs that lie on a straight line. These data are reported in table 6.

For large gain gamble pairs 3 and 4, the table indicates that 83% of 58 subjects

Table 6. Within-subject tests of betweenness

	Large gains				Small gains				Small losses			
		Fraction of choices				Fraction of choices				Fraction of choices		
Pairs	Sample size	Same	Less-more	More-less	Sample size	Same	Less-more	More-less	Sample size	Same	Less-more	More-less
Hypotenuse												
3–4	58	0.83	0.14	0.03	56	0.71*	0.13	0.16	60	0.77	0.08	0.15
4–13	25	0.64*	0.36	0.00	21	0.91	0.10	0.00	21	0.62*	0.10	0.29
7–8	97	0.65**	0.30	0.05	97	0.63**	0.31	0.06	97	0.66**	0.09	0.25
10–11	59	0.70*	0.20	0.10	58	0.78	0.16	0.07	57	0.68*	0.04	0.28
11–14	23	0.61*	0.22	0.17	22	0.68	0.27	0.05	18	0.72	0.17	0.11
Means	262	0.69**	0.24	0.07	254	0.71*	0.21	0.08	253	0.69**	0.08	0.23
Edge pairs												
2–3	57	0.84	0.02	0.14	56	0.68*	0.18	0.14	58	0.66**	0.29	0.05
5–6	99	0.83	0.06	0.11	99	0.71*	0.11	0.32	97	0.81	0.09	0.09
9–10	60	0.53**	0.03	0.43	59	0.46**	0.10	0.44	59	0.73	0.15	0.12
Means	216	0.75*	0.04	0.21	214	0.63**	0.13	0.31	214	0.75*	0.16	0.09

*$p < .05$, z-test of fraction of same choices $< .867$.
**$p < .01$.

chose the same gamble in both pairs (i.e., the more risky gamble or the less risky gamble). Fourteen per cent chose the less risky gamble in pair 3 and the more risky gamble in pair 4, suggesting concave indifference curves (quasiconvex preferences), a dislike of mixtures. Three percent chose the more risky gamble in pair 3 and the less risky gamble in pair 4, suggesting convexity (quasiconcave preferences), a preference for mixtures. If the betweenness violations are random, then the percentage of same choices (83% in pairs 3–4) should equal the percentage of same choices when the same gamble was presented twice (back to back) to test reliability, 86.7%. Asterisks denote significance levels for the test of whether the fraction of same choices was equal to 86.7%. Most of the z-statistics are significant, many at the 1% level.

If violations are random, then the fractions of less–more and more–less choices should be about equal. Generally they are not; curves appear to be concave (a larger number of less–more choices) on the hypotenuse and convex on the edges for gain gambles, and the opposite for loss gambles. (In these data, statistical tests show that a difference in the less–more and more–less fractions of more than .1 is significant at the 5% level.)

Coombs and Huang (1976) tested betweenness also. Their work was motivated by *portfolio theory*, the idea that preferences for gambles depend upon expectation and risk, and people have some optimum level of risk (i.e., risk preferences are single-peaked). According to portfolio theory, betweenness can be violated because a probability mixture of gambles A and B may have a risk level that is closer to the optimum than the risk levels of A and B are. In one set of three gamble pairs, with money amounts rounded to dollars, their subjects violated betweenness only 14% of the time (less than the 19% expected by chance[25]). In a set of gambles with unusual money amounts (like $2.13) they violated betweenness 46% of the time. Chew and Waller (1986) found 27% violations.

In Coombs and Huang's data, portfolio theory accounted for about 60% of the violations of betweenness. Portfolio theory predicts only about half of the violations in our data, and 30% of Chew and Waller's violations.[26]

Fanning out is tested similarly in table 7. Consider two gamble pairs that lie on an edge, such that the gambles in one pair stochastically dominate the gambles in the other pair (e.g., hypotenuse pairs 1 and 8). The first important statistic is the fraction of subjects who chose the same gamble in both pairs (i.e., the less risky or the more risky gamble); this is the fraction consistent with EU (60% for large gain pairs 1–8). The second important statistic is the fraction of EU violations that are explained by fanning out (75% for large gains pairs 1–8). The asterisks in the table denote the results of a z-test comparing the fraction of choices consistent with EU with 75.9% (the reliability for identical pairs separated by 0–3 other pairs, as the fanning-out pairs were), and a z-test comparing the fraction of violations consistent with fanning out with 50%.

Table 7 shows that there is little systematic fanning out or fanning in for pairs on the hypotenuse (as EURDP predicts) and on the left edge. There is strong fanning out for pairs on the lower edge (for large gain gambles) and two edge pairs (i.e., pairs on the left and lower edge).

Table 7. Within-subject tests of fanning out

Pairs	Large gains			Small gains			Small losses		
	Sample size	Fraction consistent with EU	Fraction of violations fanning out	Sample size	Fraction consistent with EU	Fraction of violations fanning out	Sample size	Fraction consistent with EU	Fraction of violations fanning out
Hypotenuse									
1–8	10	0.60	0.75	10	0.05	0.50	8	0.63	0.34
13–14	48	0.75	0.50	39	0.69	0.68	43	0.70	0.70
8–12	10	0.60	0.75	10	0.50	0.20	10	0.90	0.00
Means	68	0.71	0.60	59	0.63*	0.53	61	0.72	0.60
Left edge									
1–2	29	0.80	0.20	29	0.62	0.18*	28	0.68	0.12**
1–5	29	0.79	0.50	29	0.59*	0.67	29	0.66	0.90**
2–5	67	0.72	0.42	69	0.61*	0.70*	66	0.64*	0.67*
Means	125	0.75	0.39	127	0.61**	0.58	123	0.65*	0.61
Lower edge									
5–9	70	0.39**	0.93**	70	0.60*	0.64	28	0.61	0.54
5–12	30	0.37**	0.95**	30	0.50**	0.47	29	0.62	0.37
9–12	30	0.67	0.90**	29	0.66	0.30	28	0.61	0.54
Means	130	0.45**	0.93**	129	0.59**	0.53	85	0.61*	0.48
Two edges									
1–9	30	0.37**	0.84**	29	0.59*	0.67	28	0.57*	0.26*
1–12	30	0.47**	0.94**	29	0.59*	0.50	27	0.56*	0.67
2–9	68	0.43**	0.90**	70	0.44**	0.85**	69	0.48**	0.72**
2–12	30	0.47**	0.81**	30	0.43**	0.82**	28	0.57*	0.25*
Means	128	0.45**	0.89**	129	0.47**	0.78**	124	0.52**	0.62**

$^{*}p < .05$. See text for descriptions of tests.
$^{**}p < .01$.

Since fanning in is rarely significant, the theories that predict fanning out (lottery-dependent EU, weighted EU, and Machina's hypothesis) are roughly correct; but the theories that predict fanning in on the left edge (EURDP, prospect theory) or hypotenuse (prospect theory) are not rejected either.

One can also test fanning out with triples of gamble pairs. The results are too detailed to report, but they can be summarized briefly. For triples of gamble pairs on the hypotenuse, EU accounts for 50%, 49%, and 45% of the patterns for large gains, small gains, and small losses respectively. (If people obey EU unreliably, EU should account for 65.8% of the patterns.) For triples on edges, EU accounts for 31%, 33%, and 40%; fanning out accounts for 53%, 36%, and 15% (excluding patterns consistent with EU); and fanning in accounts for 7%, 14%, and 29%. Fanning out accounts for EU violations fairly well for gain gambles, while fanning in accounts for loss gambles.

4.3. Other evidence

Fanning out has been tested in several other ways. Much of the evidence comes from two well-known kinds of problems, involving *common ratios* and *common consequences.*

Common-ratio problems usually involve two gamble pairs, such as $(rp,X_H;1-rp,0)$ vs. $(p,X_L;1-p,0)$ and $(rpq,X_H;1-rpq,0)$ vs. $(pq,X_L;1-pq,0)$. The probabilities of the prizes X_H and X_L have a common ratio, because p/rp is the same as pq/rpq. (In our gamble pairs, a common ratio problem would compare the choice between the less risky pair 5 gamble and the more risky pair 8 gamble with the choice in gamble pair 12.) People often choose $(p,X_L;1-p,0)$ from the first pair (especially when $p = 1$) and choose $(rpq,X_H;1-rpq,0)$ from the second pair, which violates EU.

MacCrimmon and Larsson (1979) found extremely high rates of EU violation (up to 70%) in common-ratio problems. The violations were largest when $p = 1$ and q was small, and when payoffs were large. Battalio, Kagel, and Komain (1988, tables 5–6) also found many common-ratio violations, most of which could be explained by fanning out. Starmer and Sugden (1987a) controlled for regret effects resulting from statistical correlation of lotteries and still observed common-ratio effects for gain gambles. The violations in their 1987b paper were strongest when $p = .6$ and $q = 1/3$, which suggests the certainty effect when $p = 1$ is not the main force behind violations. They also found some fanning in on the left edge, and found weak effects for loss gambles (in both 1987a,b).

Our comparisons of gamble pairs are examples of common-consequence problems: the gambles in pair 1 are the same as the gambles in pair 2 (for example), except for a change in the consequence which is common to them. The fanning out and fanning in we observe are violations of EU in a broad class of common-consequence problems. MacCrimmon (1965), Moskowitz (1974), Slovic and Tversky (1974), and Kahneman and Tversky (1979) found EU violations in roughly a third of their common-consequence problems. MacCrimmon and Larsson (1979) found, as we did, that the largest number of violations occurred in gamble pairs that were far apart on the lower edge of the triangle, for large payoffs. Starmer and Sugden's (1987b) common-consequence problems showed a lot of EU violations, but were not especially supportive of any alternative theory.

The pattern of fanning out we see in table 7 is roughly the same as that observed in other recent studies. Using mostly common-ratio problems with small gains and losses, Starmer and Sugden (1987b) found fanning out on the lower edge (especially their pairs 2–3, 5–6, 7–8) and some fanning in on the left edge (their pairs 11–12) as suggested by prospect theory or EURDP. However, the patterns of fanning out were quite different when probabilities were fixed and outcomes were varied (as when one looks at the same part of the triangle in figures 9–11); this is more evidence of a dependence of probability weights on payoffs that EURDP and prospect theory do not capture (Starmer and Sugden, 1987c).

Battalio, Kagel, and Komain (1988) also found fanning out on the lower edge for losses (their figure 5a) and for gains (their table 8, gambles 9 and 14). However, they observed a lot of fanning in toward the northwest corner for gain gambles (their table 8), as prospect theory and EURDP suggest.

Chew and Waller (1986) used hypothetical three-gamble pairs with a common consequence—one each from the left edge, the lower edge, and the lower left-hand corner—and a fourth pair anchored in the lower left-hand corner, to test the fanning out and betweenness properties of weighted EU. Since they used four gambles, there were sixteen possible choice patterns. Only the light hypothesis of weighted EU, which predicts fanning out, accounts for more choices *per prediction* than EU does, but fanning out and prospect theory do fairly well.[27]

Harless (1988) used common-consequence pairs that were slightly inside the edges of the triangle. His results are quite different from ours: he found some fanning out on the left edge, fanning *in* on the lower edge, and few systematic effects on the hypotenuse. His data are rather consistent with EURDP with a risk-*preferring* (convex) $g(p)$ function. Moving gambles slightly in from the edges seems to be an important change that deserves further attention.

5. Conclusions

People often violate EU; our data are no exception. In our study, the violations were typically not random and were not affected by financial incentives.

The important empirical question is whether any theory that generalizes EU can explain the violations. The results of our test and other recent tests are decidedly mixed: Each theory can account for some of the violations, but not all.

Indifference curves seem to fan out in most portions of the triangle diagrams, and fan in some portions; the fanning-out hypothesis is sometimes violated. Lottery-dependent EU predicts uniform fanning out and concavity of indifference curves (or fanning in and convexity), but there is some of each in each triangle (as prospect theory predicts). Weighted utility can account for the degree to which fanning out varies with payoffs (which EURDP and prospect theory cannot explain), but not the extensive betweenness violations. Finally, the reflection of indifference curves between gain and loss gambles suggests that subjects value *changes* in wealth, as prospect theory assumes, rather than wealth.

The results of several other recent studies corroborate most of these findings, but not all of them. In all studies there are substantial violations of EU, but no single theory can explain the patterns of violations.

Of course, one can always improve descriptive power by combining parts of theories into hybrid theories. For instance, combining the payoff-dependent weights in weighted EU with the concavity of indifference curves in EURDP (Chew and Epstein, 1987a) produces a theory that can account for betweenness violations *and* fanning out. Letting decision weights in prospect theory depend on payoffs (cf. Luce and Narens, 1985; Hogarth and Einhorn, 1987) produces a similar theory. Generalizing theories this way is tempting, but possibly unproductive. Such theories will still be unable to explain all classes of violations, and will be cumbersome to use in generating economic theory and aiding decision makers. The simplicity of EU and fanning out are appealing in comparison.

Perhaps a more useful activity is to list known violations of EU like those shown here and ask what general psychological principles might account for these phenomena (without too much initial regard for their axiomatic foundations). Such a list will be long and depressing. This study provides evidence of payoff-dependent fanning out, both concavity and convexity of indifference curves (thus violating betweenness), framing, and reflection effects. Harless's (1988) observation that gamble pairs slightly inside the triangle edges yield patterns of choice quite different from pairs on the edges is an important puzzle. There are many other violations like preference reversals, framing and response mode effects, regret effects, and ambiguity aversion.

It is also useful to work out the implications of generalized utility theories for settings in which EU is used in some descriptive or prescriptive way, such as consumer behavior (Thaler, 1985), game theory (Crawford, in press), insurance, asset pricing (Chew and Epstein, 1987b), decision analysis, and risk management (see also Weber and Camerer, 1987, pp. 147–148). It might happen that the theories considered here are so awkward to apply clearly, or that their implications are so similar to those of EU, that a modest increase in descriptive power is too high a price to pay.

Acknowledgments

Thanks to Jon Baron, Dan Friedman, Jack Hershey, Chew Soo Hong, John Kagel, Dan Kahneman, Paul Kleindorfer, Marc Knez, Howard Kunreuther, Graham Loomes, Ken MacCrimmon, Mark Machina, Lisabeth Miller, Tom Russell, Rakesh Sarin, Martin Weber, and seminar participants at the University of Chicago Center for Decision Research, the University of Arizona, the NYU Graduate School of Management, and the Economic Science Association. These experiments were funded by the Wharton Risk and Decision Processes Center.

Notes

1. Other interesting theories include disappointment (Bell, 1985; Loomes and Sugden, 1986), dual bilinear utility (Luce and Narens, 1985), lexicographic preferences (e.g., Encarnacion, 1987), and similarity-based preferences (Rubinstein, 1986). We have not explored the predictions of these theories in this experiment. Psychologists have also studied theories in which preferences are determined by the statistical moments of gambles (see Payne, 1972). These theories cannot account for choices observed in figures 10–11 because when the mean of two gambles is the same, gambles with lower variance, kurtosis, etc. are sometimes chosen and sometimes rejected. Another class of theories suggests that people weight the dimensions of risk (usually probabilities and outcomes) linearly to determine preference. These theories cannot account for our data either, but enough puzzles remain that information-processing-based theories like these deserve renewed attention.

2. An alternative is to elicit subjects' indifference curves, using lottery equivalents (McCord and deNeufville, 1986), or to establish tight bounds on indifference curves from pairwise choices (cf. MacCrimmon and Toda, 1969). A pilot experiment suggested that the latter method is feasible, but it appears less efficient and more prone to elicitation bias (e.g., Hershey, Kunreuther, and Schoemaker, 1982) than the method used here. (Hey and Strazzera, 1988, have tried eliciting curves.)

3. MacCrimmon and Larsson (1979) found that out of 20 decision rules, transitivity was the rule subjects agreed with most strongly (a mean rating of 8.8 on an 11-point scale). Comparability ranked seventh highest (mean rating 6.9) and the common-consequence property of independence ranked lowest (mean rating 5.0).

4. Positive affine transformation yields equivalent representations because the constants a and b subtract out and divide out, respectively, from equation (1). Put differently, the origin and scale of the utility function can be arbitrarily defined.

5. Weber showed that the weighted utility of a gamble is equivalent to the expected utility of a modified gamble in which the outcome X, once it occurs, is only received with probability $W(X)$ (otherwise the gamble G is received). In an equivalent modified gamble, the outcome X_i is first imagined with probability $W(X_i)/\Sigma_{i=1}^{n} W(X_i)$. If it is imagined, X_i actually occurs with probability $p(X_i)$ (otherwise the gamble G occurs). In the latter interpretation, the number $W(X_i)/\Sigma_{i=1}^{n} W(X_i)$ might represent the fraction of psychological scenarios in which the possibility of X_i occurs.

6. The gambles used later yield zero or either of two nonzero outcomes, with respective probabilities p_L, p_M, and p_H (for positive nonzero outcomes). The less risky gamble, call it S, is usually the risky gamble, call it R, with probabilities $p_L + .1, p_M - .2$, and $p_H + .1$. The realizations of gambles correspond to consequences of acts in a state framework (cf. Savage, 1954):

State	1	2	3	4	5
Probability	p_H	.1	$p_M - .2$	.1	p_L
R consequence	X_H	X_H	X_M	0	0
S consequence	X_H	X_M	X_M	X_M	0

Notice that the gambles R and S only have different consequences in two states, which both have objective probability .1. Since $\phi(X,X) = 0$, R is preferred to S iff $.1\phi(X_H,X_M) + .1\phi(0,X_M) > 0$. Regret theory predicts that preference for R over S must be fixed as the three probabilities p_L, p_M, and p_H vary, while other generalized theories allow changes in preference.

7. It is not clear whether these properties hold in general. Sarin (private communication) indicated that similar properties were found to hold in numerical simulations, but proofs of general properties are not known.

8. If we increase both p_L and p_H by small amounts, the $1 - p_L - p_H$ terms in (12) unambiguously decrease. The term c_F will change by *relatively* little, because it is lowered by an increase in p_L and raised by an increase in p_H. The direction of change in dp_H/dp_L will therefore be determined by the decrease in $1 - p_L - p_H$.

9. The picture of a hypothetical weighting function in Kahneman and Tversky (1979) is remarkably like that derived empirically by Preston and Baratta (1948, figure 1).

10. The curves in figure 6 were calculated assuming $U(X_M) = (U(X_H) + U(X_L))/2$ and approximating $\pi(p)$ by the quadratic function $\pi(p) = .069 + .244p + .63p^2$ for $0 < p < 1$ and $\pi(1) = 1, \pi(0) = 0$. That function was calculated by fitting a quadratic function to three points that were visually estimated from the hypothetical $\pi(p)$ graph in Kahneman and Tversky (1979). Of course, figure 6 is merely illustrative; the general properties of prospect theory that we test are those embodied in dp_H/dp_L in equation (20).

11. Yaari (1987) and some others work with the *de*cumulative distribution function $1 - F(x)$. In that case, risk aversion (or preference) corresponds to con*vex*ity (con*cav*ity) of $g(q)$, the opposite of the form used in this paper.

12. For instance, Moskowitz (1974) and Keller (1985) found that matrix representations like those used here (*without* proportional representation of payoff sizes) led to fewer violations of EU than written statements or other formats. Harless (1988) found that providing subjects with expected values and variances did not affect choices.

13. The author was introduced toward the end of the first day of classes, during which some class time goes unused by tradition. He read the instructions aloud and answered questions, and then subjects made chocies silently. The questionnaires took 5-25 minutes to complete. Some common questions (and responses) were: How much do we pay for these lottery tickets? (Nothing; assume they were gifts, or punishments if the outcomes are losses, and you must choose one or the other.) Is this part of your research? What are you trying to test? (Yes. We are interested in how people make choices involving uncertain outcomes.)

14. The experimenter brought around two boxes containing tickets. The first box contained 100 tickets, with 25 of each numbered 1, 2, 3, 4. The subjects each drew a ticket, which determined one of the four gamble pairs. They then played the gamble they preferred in the chosen pair by choosing a ticket from a second box, containing 100 tickets numbered 1 through 100. That ticket determined the payoff for the gamble they preferred, in the pair that was determined by the first ticket. Subjects recorded their ticket numbers, showed them to an experimenter, and were paid.

15. To maintain credibility in the procedure through most of the experiment, the classes in which subjects were allowed to change their minds were the last two of the seven classes.

16. Indifference judgments were only allowed in one class. Three subjects said they were indifferent in almost every gamble pair and no others indicated any indifference. Battalio, Kagel, and Komain (1987) and Harless (1988) found few indifference judgments. Subjects may falsely report indifference in order to finish the task quickly, or fail to report true indifference (if they are allowed to) because they think experimenters do not want them to. Furthermore, playing chosen gambles is tricky if indifference is allowed. If you flip a coin to determine which of two indifferent gambles is played, you are implicitly assuming that preferences satisfy betweenness. (If preferences do not satisfy betweenness, subjects may misleadingly state, or fail to state, indifference because of preference or aversion to probability mixtures.)

17. When there was a gap of 0–3 pairs between presentations, there were 24.1% reversals (n = 112); a 4–7 pair gap had 33.1% reversals (n = 121); an 8–11 pair gap had 36.2% reversals (n = 115).

18. This amount of reversal (two of 80) is quite low in comparison to the one third of subjects who reversed preferences on identical gambles. Conscious reversals are apparently much rarer than unconscious reversals.

19. These data exclude 39 subjects who chose equal numbers of more and less risky choices, for either gain or loss gambles.

20. Data from all three payment conditions were pooled. Choices among loss gambles were pooled too, despite the differences between subjects who played a loss gamble and those who did not (table 3), because the *magnitude* of differences was quite small. None of the conclusions we draw are reversed if the data from the subjects who actually played a loss gamble are excluded.

21. However, if all subjects have identical indifference curves but make random errors in choices, steeper thick lines will be an indication of steeper individual indifference curves.

22. The sample sizes are between 100 and 130 for most gamble pairs (60–65 for pairs 13 and 14), so the standard errors of the fractions of less risky choice are about .04. The standard error of the *difference* of two independent fractions is about .06 (.04 times the square root of 2), so a difference of .10 is almost two standard errors.

23. The chi-squared statistics are 200, 106, and 76 for the three figures, respectively ($p < .001$ for all).

24. This violation of prospect theory can be accommodated by assuming that a .1 probability of X_H is overweighted but a .1 probability of X_M is underweighted.

25. Their subjects gave the same response about 4.5 times, on average, in five identical replications of each gamble. Assuming responses are independent and using the binomial distribution, we can estimate that the unreliability probability that generates an expected value of 4.5 identical responses is about 10%. The probability of a random violation of betweenness in responses to three gambles is then $1 - (.9)(.9)$, or 19%.

26. In our jargon, the *more–less* pattern is consistent with portfolio theory because the riskier gamble *B* in the first pair can be preferred to the less risky gamble *A* (because *A* is not risky enough), while *B* is also preferred to the more risky gamble *C* in the second pair (because *C* is too risky). By the opposite logic, the *less–more* pattern is *not* consistent with portfolio theory; but most betweenness violations are less–more choices.

27. The number of patterns predicted by each theory and the total fraction of choices explained per predicted pattern were: EU (2, 13.4%); weighted EU (light hypothesis, 4, 14.7%; heavy hypothesis, 4, 8.6%); implicit EU (8, 9.2%); fanning out (6, 11.8%); lottery-dependent EU (14, 7%; see Becker and Sarin, 1987); and prospect theory (6, 10.3%). Further details are available from the author.

Appendix: Experimental materials (played gain condition)

This is an experiment about lotteries with uncertain payoffs. You will be given a series of choices between two lotteries. For each pair of lotteries, you should indicate which of the two lotteries you prefer to play. You will actually get the chance to play one of the lotteries you chose, so you should think carefully about which lotteries you prefer.

Here is a pair of lotteries like the ones you will see: [Figure A. 1 is showing]

The outcomes of the lotteries will be determined by a random number between 01 and 100. Each number between (and including) 01 and 100 is equally likely to occur. In the example above, the left lottery, labeled *A*, pays nothing (0) if the random number is between 01 and 40. Lottery A pays five dollars ($5) if the random number is between 41 and 100. Notice that the picture is drawn so that the height of the line between 01 and 40 is 40% of the distance from 01 to 100. The rectangle around $5 is 60% of the distance from 01 to 100.

In the example above, the lottery on the right, labeled *B*, pays nothing (0) if the random number is between 01 and 50, five dollars ($5) if the random number is between 51 and 90, and ten dollars ($10) if the random number is between 91 and 100.

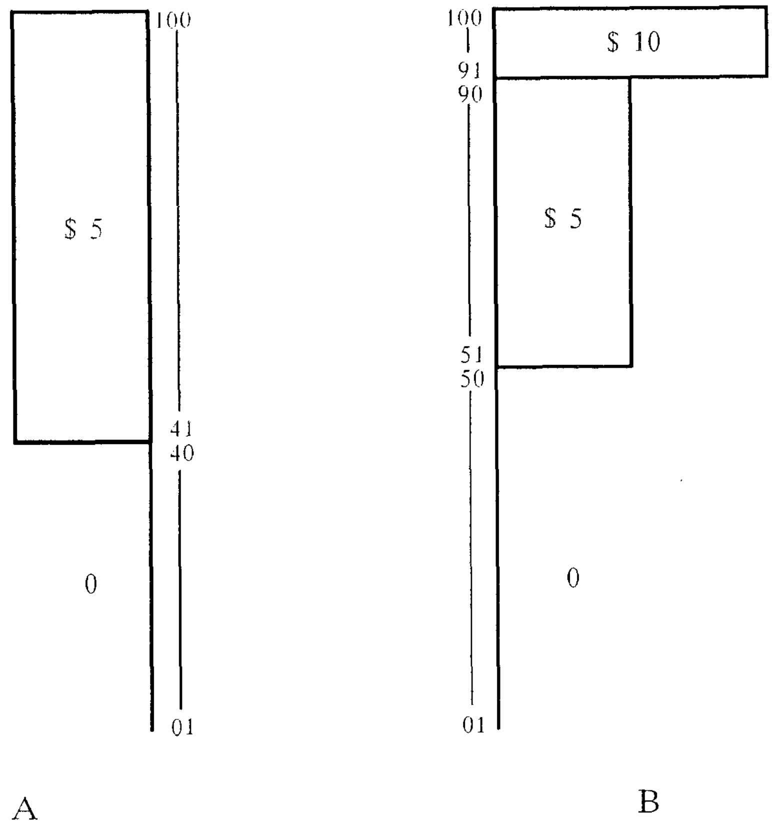

Fig. A.1. Gamble pair 9 as presented to subjects.

As with lottery A, the *heights* of the lines in lottery B represent the fraction of the possible numbers that yield each payoff. For example, the height of the $10 rectangle is 10% of the way from 01 to 100. The *widths* of the rectangles are proportional to the size of their payoffs. In lottery B, for example, the $10 rectangle is twice as wide as the $5 rectangle.

Some of the lotteries involve large payoffs ($10,000 or $25,000), some involve smaller payoffs ($5 or $10), and some involve losses (−$5 or −$10). The foundation that is sponsoring this research cannot afford to pay the largest payoffs, but you will get to play one of the lotteries with smaller payoffs.

Each pair of lotteries is on a separate page. On each page, you should indicate which of the lotteries you prefer to play by circling either A, if you prefer the A lottery, or B, if you prefer the B lottery. You should approach each pair of lotteries as if it is the only pair of lotteries you are considering, because you are only going to play one of the many lotteries.

After you have worked through all the pairs of lotteries, raise your hand and an experimenter will bring you two containers of cardboard tickets. You will select one ticket from each of the two containers. One container has 100 tickets numbered 1 through 4 (25 numbered 1, 25 numbered 2, etc.); the other container has 100 tickets number 01 through 100. (If you wish, you may examine the whole containers of tickets after the experiment.)

The first ticket determines which pair of lotteries has been chosen. Some of the pages you have are numbered 1 through 4 in the upper right hand corners; the first ticket you choose determines which of those pages you will play. If the first ticket you chose is numbered 3, for example, then you will play whichever lottery you picked on page number 3.

The second ticket is the random number which determines the outcome of the lottery you chose, on the page determined by the first ticket. For instance, suppose you picked the A lottery on the first page of these instructions. If the random number was 37, you would win nothing; if it was 93, you would get $5. If you picked the B lottery and drew the number 37, you would get nothing; if it was 93, you would get $10.

Therefore, your payoff is determined by three things: by which page is chosen, as determined by the first ticket you chose; by which lottery you picked on that page; and by the outcome of that lottery. This procedure is explained again on the last page.

This is not a test of whether you can pick the best lottery in each pair, because none of the lotteries are necessarily better than the others. Which lotteries you prefer is a matter of personal taste. The people next to you will have different lotteries, and may have different tastes, so their responses should not matter to you. Please work silently, and make your choices by thinking carefully about each lottery. Any changes or erasures will make you ineligible for a payoff.

After you have chosen one lottery on each page, raise your hand and the experimenters will come around with the containers of tickets. After you have chosen your tickets and determined your payoff, sign the receipt on the last page, give your questionnaire to one of the experimenters, and you will be given your payoff in cash. Then you are free to leave. Feel free to ask questions now, or during the experiment.

You have now finished making choices. Follow the instructions below in numerical order.

[last pages]

1) Please check to be sure that you have made one choice (by circling either A or B) on each page.

2) Raise your hand. The experimenters will bring the two containers of tickets.

3) Draw one ticket from each container, then write their numbers below.

First ticket
(1 through 4): __________

Second ticket
(01 through 100): __________

4) Turn back to
the page with this number in the corner.
5) For the lottery you chose, write the payoff for this number
here: __________

This is your payoff for the experiment.
6) Sign the receipt below and bring this questionnaire to an experimenter. He will give you your payoff, or collect it from you if it is negative. Be sure to return your tickets to the experimenter.

Name ______________________________ Received __________
Social Security No. ______________________________ Date __________

References

Allais, Maurice. (1953). "Le Comportement de L'homme Rationel Devant le Risque, Critique des Postulates et Axiomes de L'ecole Americaine." *Econometrica* 21, 503–546.

Allais, Maurice. (1979). "The So-called Allais Paradox and Rational Decisions Under Uncertainty". In Maurice Allais and O. Hagen (eds.), *The Expected Utility Hypothesis and the Allais Paradox,* Dordrecht: D. Reidel Publishing Company.

Battalio, Ray C., John H. Kagel, and Don N. MacDonald. (1988). "Animals' Choices Over Uncertain Outcomes: Some Initial Experimental Evidence," *American Economic Review* 75, 597–613.

Battalio, Ray C., John H. Kagel, and Jiranyakul Komain. (1988). "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results." Texas A&M Department of Economics working paper.

Becker, Joao L., and Rakesh Sarin. (1987). "Lottery Dependent Utility," *Management Science* 33, 1367–1382.

Bell, David. (1982). "Regret in Decision Making under Uncertainty," *Operations Research* 30, 961–981.

Bell, David. (1985). "Disappointment in Decision Making Under Uncertainty," *Operations Research* 33, 1–27.

Chew, Soo Hong. (1983). "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica* 51, 1065–1092.

Chew, Soo Hong. (1984). "An Axiomatization of the Rank Dependent Quasilinear Mean Generalizing

the Gini Mean and the Quasilinear Mean," John Hopkins University Department of Political Economy working paper.
Chew, Soo Hong. (1985). "Implicit-weighted and Semi-weighted Utility Theories, M-estimators, and Non-demand Revelation of Second-price Auctions for an Uncertain Auctioned Object," Johns Hopkins University Department of Political Economy working paper #155.
Chew, Soo Hong, and Kenneth R. MacCrimmon. (1979). "Alpha-nu Choice Theory: An Axiomatization of Expected Utility," University of British Columbia Faculty of Commerce working paper #669.
Chew, Soo Hong, and Larry G. Epstein. (1987a). "A Unifying Approach to Axiomatic Non-Expected Utility Theories," John Hopkins University Department of Political Economy working paper.
Chew, Soo Hong, and Larry G. Epstein. (1987b). "Non-Expected Utility Preferences in a Temporal Framework with an Application to Consumption-Savings Behavior," John Hopkins University Department of Political Economy working paper.
Chew, Soo Hong, and William S. Waller. (1986). "Empirical Tests of Weighted Utility Theory," *Journal of Mathematical Psychology* 30, 55–72.
Coombs, Clyde, and Lily Huang. (1976). "Tests of the Betweenness Property of Expected Utility," *Journal of Mathematical Psychology* 13, 323–337.
Crawford, Vincent P. (forthcoming). "Equilibrium without Independence," *Journal of Economic Theory.*
Dekel, Eddie. (1986). "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory* 40, 304–318.
Edwards, Ward. (1954a). "The Theory of Decision Making," *Psychological Bulletin* 51, 380–417.
Edwards, Ward. (1954b). "The Reliability of Probability Preferences," *American Journal of Psychology* 67, 68–95.
Edwards, Ward. (1955). "The Prediction of Decisions Among Bets," *Journal of Experimental Psychology* 51, 201–214.
Einhorn, Hillel J., and Robin M. Hogarth. (1985). "Ambiguity and Uncertainty in Probabilistic Inference," *Psychological Review* 92, 433–461.
Ellsberg, Daniel. (1961). "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75, 643–669.
Encarnacion, J. (1987). "Preference Paradoxes and Lexicographic Choice," *Journal of Economic Behavior and Organization* 8, 231–248.
Fishburn, Peter. (1982). "Nontransitive Measurable Utility," *Journal of Mathematical Psychology* 26, 31–67.
Fishburn, Peter. (1983). "Transitive Measurable Utility," *Journal of Economic Theory* 31, 293–317.
Fishburn, Peter. (1984). "SSB Utility Theory: An Economic Perspective," *Mathematical Social Science* 8, 63–94.
Grether, David. (1981). "Financial Incentive Effects and Individual Decision Making", California Institute of Technology working paper 401.
Grether, David, and Charles R. Plott. (1979). "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69, 623–638.
Handa, Jagdish. (1977). "Risk, Probabilities, and a New Theory of Cardinal Utility," *Journal of Political Economy* 85, 97–122.
Harless, David W. (1988). "Predictions about Indifference Curves in the Unit Triangle: A Test of Competing Decision Theories", Grinnell College working paper.
Hershey, John, Howard Kunreuther, and Paul Schoemaker. (1982). "Sources of Bias in Assessment Procedures for Utility Function," *Management Science* 28, 936–954.
Hey, John D. (1984). "The Economics of Optimism and Pessimism: A Definition and Some Applications," *Kyklos* 37, 181–205.
Hey, John D. and Elisabetta Strazzera. (1988). "Estimation of Indifference Curves in the Marschak-Machina Triangle," University of York Department of Economics working paper.
Hogarth, Robin, and Hillel Einhorn. (1987). "Venture Theory: A Model of Decision Weights," University of Chicago Center for Decision Research working paper.
Hogarth, Robin, and Howard Kunreuther. (forthcoming). "Risk, Ambiguity, and Insurance," *Journal of*

Risk and Uncertainty.

Kahneman, Daniel, and Amos Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263–291.

Karmarkar, Uday S. (1978). "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model," *Organizational Behavior and Human Performance* 21, 61–72.

Keller, Robin L. (1985). "The Effects of Problem Representation on the Sure-Thing and Substitution Principles," *Management Science* 31, 738–751.

Loomes, Graham. (1988). "Further Evidence of the Impact of Regret and Disappointment in Choice under Uncertainty," *Economica* 55, 47-62.

Loomes, Graham, and Robert Sugden. (1982). "Regret theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal* 92, 805–824.

Loomes, Graham, and Robert Sugden. (1986). "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies* 53, 271–282.

Loomes, Graham, and Robert Sugden. (1987a). "Some Implications of a More General Form of Regret Theory," *Journal of Economic Theory* 41, 270–287.

Loomes, Graham, and Robert Sugden. (1987b). "Testing for Regret and Disappointment in Choice under Uncertainty," *Economic Journal* 97, 118–129.

Luce, R. Duncan, and Louis Narens. (1985). "Classification of Concatenation Measurement Structures According to Scale Type," *Journal of Mathematical Psychology* 29, 1–72.

MacCrimmon, Kenneth R. (1965). "An Experimental Study of the Decision Making Behavior of Business Executives," unpublished dissertation, University of California, Los Angeles.

MacCrimmon, Kenneth R., and Stig Larsson. (1979). "Utility Theory: Axioms Versus Paradoxes." In Maurice Allais and O. Hagen (eds.), *The Expected Utility Hypothesis and the Allais Paradox.* Dordrecht: D. Reidel Publishing Co..

MacCrimmon, Kenneth R., and Masanao Toda. (1969). "The Experimental Determination of Indifference Curves," *Review of Economic Studies* 36, 433–451.

Machina, Mark. (1982). "Expected Utility Analysis Without the Independence Axiom," *Econometrica* 50, 277–323.

Machina, Mark. (1983). "The Economic Theory of Individual Behavior Toward Risk," Technical Report #433, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Markowitz, Harry. (1952). "The Utility of Wealth," *Journal of Political Economy* 60, 151–158.

Marschak, Jacob. (1950). "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica* 18, 111–141.

McCord, Mark, and Richard de Neufville. (1986). "Lottery Equivalents: Reduction of the Certainty Effect Problem in Utility Assessment," *Management Science* 32, 56–60.

Moskowitz, Herbert. (1974). "Effects of Problem Representation and Feedback on Rational Behavior in Allais and Morlat-type Problems," *Decision Sciences* 5, 225–242.

Payne, John W. (1973). "Alternative Approaches to Decision Making Under Risk: Moments versus Risk Dimensions," *Psychological Bulletin* 80, 439–453.

Preston, Malcolm G., and Philip Baratta. (1948). "An Experimental Study of the Auction-Value of an Uncertain Outcome", *American Journal of Psychology* 61, 183–193.

Quiggin, John. (1982). "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization* 3, 323–343.

Quiggin, John. (1985). "Subjective Utility, Anticipated Utility, and the Allais Paradox," *Organizational Behavior and Human Decision Processes* 35, 94–101.

Roell, Alissa. (1987). "Risk Aversion in Quiggin and Yaari's Rank-Order Model of Choice under Uncertainty," *Economic Journal* 97, 143–159.

Rubinstein, Ariel. (1986). "Similarity and Decision Making Under Risk (Is there a Utility Theory Resolution to the Allais Paradox?)," Hebrew University Department of Economics research report no. 156.

Schoemaker, Paul. (1982). "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations," *Journal of Economic Literature* 20, 529–563.

Segal, Uzi. (forthcoming). "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review.*

Slovic, Paul, and Sarah Lichtenstein. (1983). "Preference Reversals: A Broader Perspective," *American Economic Review* 73, 596–605.

Slovic, Paul, and Amos Tversky. (1974). "Who Accepts Savage's Axiom?," *Behavioral Science* 19, 368–373.

Smith, Vernon L. (1982). "Microeconomic Systems as an Experimental Science," *American Economic Review* 72, 923–955.

Starmer, C., and Robert Sugden. (1987a). "Experimental Evidence of the Impact of Regret on Choice Under Uncertainty," Economics Research Centre discussion paper no. 23, University of East Anglia.

Starmer, C., and Robert Sugden. (1987b). 'Violations of the Independence Axiom: An Experimental Test of Some Competing Hypotheses," Economics Research Centre discussion paper no. 24, University of East Anglia.

Starmer, C., and Robert Sugden. (1987c). "Testing Prospect Theory," Economics Research Centre discussion paper no. 26, University of East Anglia.

Sugden, Robert. (1986). "New Developments in the Theory of Choice Under Uncertainty," *Bulletin of Economic Research* 38, 1–24.

Thaler, Richard. (1985). "Mental Accounting and Consumer Choice," *Marketing Science* 4, 199–214.

Tversky, Amos, and Daniel Kahneman. (1987). "Rational Choice and the Framing of Decisions," *Journal of Business* 59, S251–S278. Reprinted in Robin Hogarth and M. Reder (eds.) *Rational Choice: The Contrast Between Economics and Psychology,* Chicago: University of Chicago Press.

Weber, Martin, and Colin F. Camerer. (1987). "Recent Developments in Modelling Preferences Under Risk," *OR Spektrum* 9, 129–151.

Weber, Robert J. (1982). "The Allais Paradox, Dutch Auctions, and Alpha-Utility Theory," Northwestern University MEDS Department discussion paper #536.

Yaari, Menahem E. (1987). "The Dual Theory of Choice Under Risk," *Econometrica* 55, 95–115.