ECO 317 – Economics of Uncertainty – Fall Term 2009
Notes for lectures
10. Critiques of Expected Utility Theory

From its earliest days, expected utility theory met several criticisms. Some were based on a priori arguments that its underlying assumptions were unreasonable, some were based on experimental or empirical evidence that behavior did not conform to its predictions, and some combined the two lines of criticism. Numerous alternative theories were also offered; some of them met similar or weightier objections in their turn. Here we review just a few prominent critiques and alternatives, and then a brief overview.

1. Some Critiques

Reference Points, Status Quo Effects, and Framing

This is perhaps the most fundamental critique, not merely of expected utility theory, but of much standard economic theory of optimal decision-making. That theory stipulates an objective function depending only on "final" positions – the quantities of goods and services actually consumed (including extensions to choices over time, where commodities are labeled by the date when they are consumed, and to uncertainty, where they are labeled by the state of the world in which they are consumed). Psychologists have argued that people's perception and evaluation of outcomes is importantly affected by a *reference position,* which may be the same person's previous situation (the *status quo*), or something to which his attention is drawn by the way the issue is *framed.* (Daniel Kahneman, *American Economic Review,* December 2003, especially pp. 1454-60.) This is supported by recent research in *neuroeconomics,* which monitors people's brain activity while they are making economic decisions. Roughly speaking, the finding is that gains and losses, or benefits and costs, are processed by different parts of the brain and the calculations are not perfectly brought together by any superior process.

For economic choices without uncertainty, this effect was most dramatically demonstrated by Richard Thaler's experiments on the *endowment effect.* Essentially, he sold small objects like coffee mugs to his laboratory subjects, and then offered to buy them back. He found that people would part with the objects only for premia as high as 100 percent over the prices they were willing to pay when buying. The value of the good appeared higher when it was viewed as something that could be lost than when it was viewed as something to be acquired.

However, John List performed similar experiments with subjects who had experience of trading sports cards, and found that they were less subject to the endowment effect, especially if their experience was on the selling side. They seemed to have learned from experience to base choices on some underlying or long-term value, and not on the emotions associated with acquiring or losing things.

The general issue of how far actual behavior departs from the rationality assumed in conventional economic theory – complete and transitive preferences defined over the final position – is still unsettled, and the answer is likely to depend on the context in which the choice is being made. The answers that people have found have also varied with the methods of research employed. Broadly speaking, the distinction is between *laboratory experiments* on one hand, and *field experiments* or *observations of real-world phenomena* on the other hand. Laboratory settings allow scientific control; one can vary just the condition whose effect one wants to study. In the field or the real world, too many uncontrolled things are changing. At best one can make allowance for them using statistical methods such as instrumental variables, and the choice and validity of these methods is often a matter of judgment beyond strict scientific criteria. However, laboratory settings and behavior can be artificial. The subjects are often college students with limited wealth and new to the kind of choice that is presented; the stakes are sometimes so small that the subjects may not take them seriously. Extrapolating from these observations to behavior of people who are regular participants in their economic activities and trades, have experience in making the same kinds of decisions, and have large amounts or even livelihoods at stake, is itself problematic. Thus each of the methods has some merits and some defects. I believe that we should take all these findings seriously, without being totally convinced by any one, and continue to develop different theoretical and empirical methods to elaborate and test all the hypotheses. More likely than not, most of the arguments and alternatives will find some range of applicability, but none will command universal validity.

Loss Aversion, First-Order Risk Aversion

The endowment effect was a kind of discontinuity in valuations around the reference point. A similar phenomenon in the context of uncertainty is loss aversion, or first-order risk aversion. Figure 1 shows this in two ways.

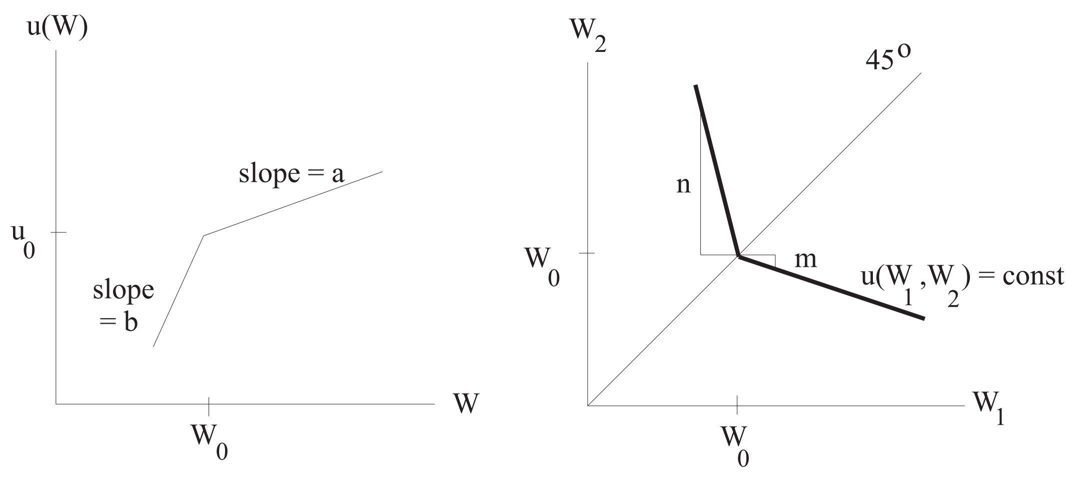

Figure 1: Loss-aversion and first-order risk aversion

The left hand panel shows a utility-of-consequences function with a discontinuous slope at the initial or status quo wealth W_0. A gain in wealth above W_0 is valued less than an equal loss below W_0. Write $u_0 = u(W_0)$ and suppose the utility-of-consequences function is

$$u(W) = \begin{cases} u_0 + a\ (W - W_0) & \text{if } W > W_0 \\ u_0 - b\ (W_0 - W) & \text{if } W_0 > W \end{cases}$$

where $b > a > 0$. Suppose this person faces a risk of gaining or losing k with equal probabilities; thus k is the standard deviation of the risk. The expected utility is

$$\tfrac{1}{2}\ (u_0 + a\ k) + \tfrac{1}{2}\ (u_0 - b\ k) = u_0 - \tfrac{1}{2}\,(b - a)\ k < u_0\,.$$

Therefore the certainty equivalent of this gamble must be less than W_0. Suppose it is $W_0 - h$, so h is the risk premium the person would be willing to pay to avoid the risk. We have

$$u(W_0 - h) = u_0 - b\ h\,.$$

Setting this equal to the expected utility of the gamble and solving for h, we have

$$h = \tfrac{1}{2}\,\frac{b - a}{b}\ k\,.$$

So the risk premium is proportional to the standard deviation of the risk. When the utility function was twice-differentiable at W_0, we found that the risk premium for small risks was proportional to the variance, or the square of the standard deviation (Handout 3, p. 6). So a loss-averse person is an order of magnitude more averse to small risks.

Two remarks on this: [1] To make the point most simply, I have taken the utility function to be linear on each side of W_0. This implies risk-neutrality for random prospects whose support lies entirely to one side or entirely to the other side of W_0. That is not essential; one could have a concave (or convex) utility function in each region. Indeed the more general prospect theory of Kahneman and Tversky does exactly that. What is important for loss aversion per se is the discontinuous change in slope (derivative) at W_0. [2] A utility-of-consequences function with a kink could be thought of as just a special case of expected utility theory, if the kink occurs at a fixed exogenous W_0. But if the kink always occurs at the status quo point, it shifts as the status quo changes. So for example if the person wins this gamble and his initial wealth next period is $W_0 + k$, then he will have a new utility-of-consequences function with a kink at $W_0 + k$. [3] This opens up some new possibilities. Does the decision-maker recognize the shifting kink and make today's choices to take into account that the immediate outcomes will affect future preferences and behavior, or does he make today's choice without such forward-looking thinking? And is it even possible to manipulate oneself into believing that the kink is now at a new point?

The right hand panel of Figure 1 shows the state-space approach, with two states and the state-contingent wealth amounts W_1 and W_2 on the axes. The indifference curves are the contours of constant utility $U(W_1, W_2)$ say, but this does NOT have to have the expected utility form.[1] I show just one indifference curve, namely the one through the no-risk status

[1] Kinks in indifference curves can exist even without any uncertainty, and the endowment effect can be interpreted in this way.

quo point (W_0, W_0). The curve has a kink at that point. Suppose the slope below the 45-degree line is m and that above the 45-degree line is n in absolute value, with $n > m > 0$. Then the points $(W_0 + 1, W_0 - m)$ and $(W_0 - 1, W_0 + n)$ are both indifferent to the status quo. So the person is indifferent between the status quo and a 50:50 gamble between these two points. But the monetary equivalent of the gamble is $(W_0, W_0 + \frac{1}{2}(n - m))$, which is first-order better than the status quo or the gamble that is indifferent to it.

The case for first-order risk aversion is made most dramatically by Rabin (*Econometrica,* September 2000). He begins with the observation that subjects in laboratory experiments often reject gambles that involve losing \$100 and winning \$110 with equal probabilities. Suppose we take this as a universal fact, valid at all initial wealth levels. Its implications can be traced most simply using a CARA utility function, although the same conclusions can be derived using an arbitrary utility-of-consequences function and comparing finite differences. We have

$$\tfrac{1}{2}\left[1 - e^{-\alpha(W_0-100)}\right] + \tfrac{1}{2}\left[1 - e^{-\alpha(W_0+110)}\right] < 1 - e^{-\alpha W_0},$$

or

$$e^{100\,\alpha} + e^{-110\,\alpha} > 2\,.$$

This turns out to imply $\alpha > 0.0009084$.

Consider such a person with $\alpha = 0.001$, say, and offer him a 50:50 chance of losing \$1000 and winning G. He will turn down this gamble if

$$\tfrac{1}{2}\left[1 - e^{-0.001\,(W_0-1000)}\right] + \tfrac{1}{2}\left[1 - e^{-0.001\,(W_0+G)}\right] < 1 - e^{-\alpha W_0},$$

or

$$e + e^{-0.001\,G} > 2\,.$$

But $e > 2$, so this is true for all G, no matter how large! So the person will reject a 50:50 gamble of losing \$1000 and winning any finite amount, no matter how large. That does not seem reasonable

Rabin also finds that such a person, faced with a portfolio choice between two assets:

[1] a riskless bond paying $\frac{1}{2}$% annual interest (\$100 becomes \$100.50), and

[2] a stock that has a mean return of 6.4% and a standard deviation of 20%, which are realistic values,

will invest only \$1639 in the stock and all the rest in the bond, no matter how large his initial wealth. He argues that both these implications are unrealistic. He offers many similar examples and calculations, and also a general algebraic argument. Together they constitute what he calls a *calibration theorem.* The idea is that by using revealed behavior in laboratory experiments to calibrate utility-of-consequences functions, we see that expected utility theory cannot explain some risk attitudes.

Others have turned this argument around, and claimed that since in the real world people do maintain portfolios that are less attractive than the gambles that are rejected by laboratory subjects, the implication should be that laboratory experiments fail to predict actual behavior, not that expected utility theory fails to predict actual behavior. Kahneman and Tversky also suggest that the kink at the reference point in their valuation function (see

Figure 3 below) is smaller when the initial wealth level is higher, so Rabin's assumption that the rejection of small, slightly better than fair, gambles in the laboratory will remain valid at all levels of wealth may not be realistic. (His subjects were the usual college students, and extrapolating from their behavior to all initial wealth levels may be invalid.) Once again, this is not a settled issue.

The Allais Paradox – Fanning Out

We saw this paradox earlier, and also saw why such behavior is incompatible with expected utility theory. (Questions 5 and 6 in the first day's questionnaire, and Precept Week 2 Question 1.) Table 1 uses different numbers to make the point even more dramatically:

	Prizes ($ millions)		
Lottery	0	1	5
A	0.00	1.00	0.00
B	0.01	0.89	0.10
C	0.89	0.11	0.00
D	0.90	0.00	0.10

Table 1: Allais paradox

Faced with a choice between lotteries A and B, many people would choose A; the 1% risk of getting nothing and thereby "losing" the $1 million looms large in their thinking. But faced with the choice between lotteries C and D, many of the same people would choose D; the chance of winning $5 million instead of $1 million would loom large, and the 1% additional probability of ending up with nothing would seem small since it was already 89% anyway. But expected utility criteria are linear in probabilities; the marginal effect of an additional 1% probability is the same whether you start at 0 or at 89%. Figure 2 shows this in the probability triangle diagram, which shows the probabilities somewhat exaggerated for visual clarity but without affecting the essential point: the four lotteries form a parallelogram.

Expected utility theory requires that the indifference map in this triangle should be a family of parallel straight lines. If the indifference lines are all flatter than the common slope of AB and CD, then the person should prefer B to A and D to C; if the indifference lines are all steeper than the common slope of AB and CD, then the person should prefer A to B and C to D.

To put the matter slightly differently, we can change lottery A into B by shifting 10% of probability from $1 million to $5 million and 1% of probability from $1 million to 0; we can change lottery C into D by an identical operation. Since expected utilities are linear in probabilities, the two operations should change expected utilities by identical amounts. Therefore going from A to B should raise expected utility if, and only if, going from C to D raises expected utility.

(Incidentally, we know that steeper lines correspond to greater risk-aversion; this fits the picture since it implies that a person with low risk aversion would prefer the riskier prospect

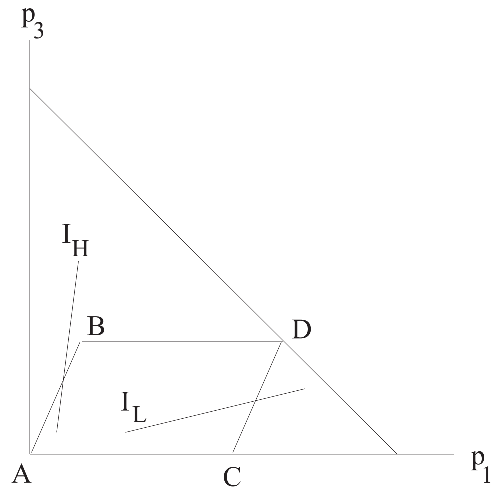

Figure 2: The Allais Paradox

B with the higher expected value of \$1.39 million to the safe \$1 million in A, and a very risk-averse person would prefer the certainty of A to the risk of B.)

To explain the commonly observed behavior in the Allais paradox, we need the indifference curves to be flatter than CD in that part of the triangle, and steeper than AB in that part. So the indifference curves should show *fanning out* from the origin. In terms of attitudes to risk, risk-aversion should be high close to the vertical axis (where the probability of the worst outcome 1 is small) and risk-aversion should be small close to the horizontal axis (where the probability of the best outcome 3 is small).

The Allais paradox has not escaped counterarguments. When it was first proposed by Allais, many adherents of expected utility argued that it was simply a failure to understand the nature of the choice being made. The Independence Axiom that underlies expected utility theory was seen as a simple "Pareto improvement" argument: If one of the lotteries in a compound lottery is replaced by a better one, the new compound lottery should rank higher in the person's preferences. The claim was that once this was properly explained, behavior would conform to expected utility theory. However, this was found not to be the case, and it is now broadly accepted that behavior conforming to the Allais paradox is perfectly consistent with the basic axioms of rationality, namely completeness and transitivity.

The next line of criticism says that paradoxical behavior occurs only for extremes of probabilities. There may be something to this; how would you choose if the probabilities in Table 1 were replaced by those in Table 2 below? Many more people are likely to show behavior conforming to expected utility theory. But the shifts in probabilities are exactly of the same kind as before. The four lotteries still form a parallelogram in the probability triangle diagram, albeit located more centrally in the triangle than close to the axes. (Useful exercise: verify this; also verify that the shifts of probability involved in going from A to B are the same as those involved in going from C to D.)

However, this counterargument does not quite rescue expected utility theory. For one thing, many actual choices do involve extreme probabilities; for example default risks on

Lottery	Prizes ($ millions) 0	1	5
A	0.30	0.40	0.30
B	0.35	0.30	0.35
C	0.40	0.30	0.30
D	0.45	0.20	0.35

Table 2: Allais paradox: 2

many bonds are quite small. And more basically, the counterargument only reinforces the basic Kahneman-Tversky argument that people's perceptions are importantly affected by their reference point.

The general idea that equal changes in probabilities are evaluated differently starting from different initial probabilities entails an overall utility function on random prospects that is not linear in probabilities. We will outline some such alternatives below.

The Ellsberg Paradox – Ambiguity Aversion

Daniel Ellsberg (who later became famous when he leaked the Pentagon Papers on the Vietnam war to the New York Times) proposed the following paradox. There are two urns:

Urn X has 50 red balls and 50 green balls.

Urn Y has 100 total balls, some red and the rest green, but the numbers of each are unknown.

A person is offered a choice between lotteries A and B as follows:

A: A ball is drawn at random from X; you get $10 if red, 0 if green.

B: A ball is drawn at random from Y; you get $10 if red, 0 if green.

Many people faced with this choice prefer lottery A. Now consider lotteries C and D:

C: A ball is drawn at random from X; you get $10 if green, 0 if red.

D: A ball is drawn at random from Y; you get $10 if green, 0 if red.

Many people faced with this choice prefer lottery C.

More remarkably, when offered both choices in different questions, say Question 1 for choice between A and B, and Question 2 for choice between C and D, many people choose A in Question 1 and C in Question 2. This raises a more basic issue about their behavior: such a choice is simply inconsistent with *any* subjective beliefs about the composition of red and green balls in urn Y.

Because urn Y has an unknown composition, there are no objective probabilities. Suppose your subjective probability of drawing a red ball from Y is p, and therefore that of drawing a green ball from Y is $(1-p)$. Since the two lotteries in each comparison have the same prizes, a rational person who prefers more to less should prefer the lottery that offers a higher probability of winning the prize. This is irrespective of whether the preferences can be represented by expected utility; it is simply a matter of monotonicity. So such a person should prefer lottery A to lottery B if $p < \frac{1}{2}$, and should prefer lottery C to lottery D if $(1-p) < \frac{1}{2}$, that is, if $p > \frac{1}{2}$. The two cannot be true at the same time; the observed preferences violate the conventional theory of choice in an even more fundamental way than the other paradoxes above.

The behavior suggests *ambiguity aversion.* The probability of drawing a red ball from urn X is unknown. Let us suppose it could take values p_1, p_2, $\ldots$ p_n, and suppose the decision-maker attaches subjective probabilities q_1, q_2, $\ldots$ q_n to these possibilities. Standard theory would regard this as a compound lottery, and reduce it to a simple lottery where the probability of drawing a red ball is $p_1\, q_1 + p_2\, q_2 + \ldots + p_n\, q_n$. But the Ellsberg paradox suggests that the reduction is not valid; people dislike the added uncertainty about the risk (ambiguity). In Donald Rumsfeld's terminology, people seem to prefer *known unknowns* to *unknown unknowns.*

2. Some Alternatives

Prospect Theory

Following Kahneman and Tversky, *Econometrica* March 1979, this has become the best supported alternative to expected utility theory. It uses an objective function of the form

$$\sum_i \pi(p_i)\ v(x_i)\,, \tag{1}$$

where i indexes the possible outcomes, x_i are the changes in outcomes (typically wealth or income or consumption quantities) relative to a reference point, p_i are the probabilities, $v(x)$ is a valuation function, and $\pi(p)$ is a weighting function. The procedure for obtaining this objective function consists of several steps:

[1] Determine the reference point relative to which gains $x > 0$ or losses $x < 0$ are measured. This is highly context-dependent; it could be an actual status quo, or it could arise because of the way the choice is framed.

[2] Find the valuation function $v(x)$ defined on deviations from the reference point, not on the final outcomes. The valuation function is generally likely to be concave for gains and convex for losses, with a kink at the reference point, as shown in Figure 3.

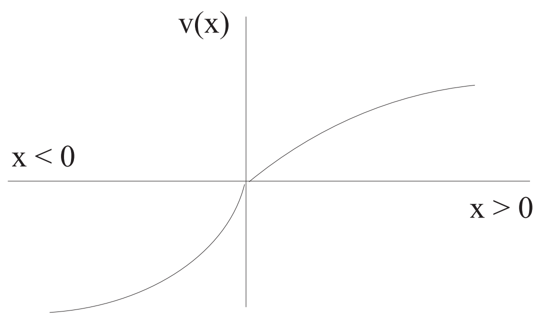

Figure 3: Prospect theory: typical valuation function

[3] Find the weighting function $\pi(p)$. These are decision weights related to probabilities, but not themselves probabilities and are not necessarily measures of degree of belief. Kah-

neman and Tversky propose a typical weighting function like the one shown in Figure 4, but others have proposed different forms.

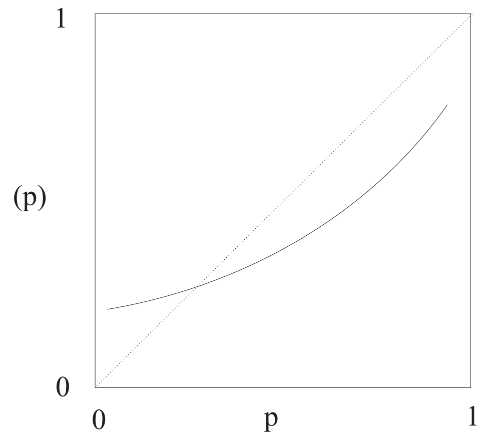

Figure 4: Prospect theory: typical weighting function

An objective function that is nonlinear in probabilities relates closely to both the Allais and Ellsberg paradoxes, and there are several variants of the idea.

Regret Theory

In Handout 3 we introduced the consequence function F, expressing the outcome c as a function of the action a and the state of the world s, namely $c = F(a, s)$. Assume that the decision-maker has complete and transitive preferences over consequences, represented by the utility function u. If he were allowed to choose his action from the feasible set A after knowing the state, he would choose it optimally to get the best outcome

$$C^*(s) = \max_{a \in A} F(a, s) \,,$$

and achieve utility $u(C^*(s))$. But he has to choose his action before knowing the state. Suppose he chooses a, and then state s materializes. His *regret* is defined as

$$R(a, s) = u(C^*(s)) - u(F(a, s)\,) \,.$$

This is intuitive: the regret from having taken action a and then seen state s materialize is the shortfall of utility below what it might have been if one could have made the best choice for that state.

Using the probabilities $Pr(s)$ of the states defined over the space of all possible states S, we can calculate *expected regret* of action a as

$$\sum_{s \in S} Pr(s)\; R(a, s) = \sum_{s \in S} Pr(s)\, u(C^*(s)) - \sum_{s \in S} Pr(s)\, u(F(a, s)) \,.$$

The first sum on the right hand side is independent of a, and the second is just expected utility. Therefore minimizing expected regret is equivalent to maximizing expected utility.

Incidentally, minimized expected regret can be viewed from a different perspective, as the additional value that would be attainable if one were able to postpone the decision until after the realization of the outcome. In other words, it is an *option value.* We will develop this idea later in the course.

But there can be other criteria based on regret. For example, one may just want to minimize the maximum possible regret, that is, choose a to minimize

$$\max_{s \in S} \; R(a, s) \,.$$

This is a rather pessimistic criterion, acting as if nature were an active player working against you - after you choose the action, nature chooses the state that would make you regret that action the most. Then your regret-minimizing choice is like a maxi-min strategy in a zero-sum game.

A more general version of this theory works for binary choices. Start with a "satisfaction function" $S(x, y)$ defined over pairs of outcomes x, y. It measures the satisfaction the decision-maker gets from outcome x when the other choice would have led to outcome y (dissatisfaction if negative). It is anti-symmetric: $S(y, x) = -S(x, y)$. Now consider choosing between two independent lotteries:

$$\begin{aligned} L^a &= (C_1,\, C_2,\, \ldots\, C_n;\; p_1^a,\, p_2^a,\, \ldots\, p_n^a) \\ L^b &= (C_1,\, C_2,\, \ldots\, C_n;\; p_1^b,\, p_2^b,\, \ldots\, p_n^b) \,. \end{aligned}$$

Now calculate the expected net satisfaction from choosing L^a instead of L^b:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i^a \; p_j^b \; S(C_i, C_j) \,.$$

Choose L^a if this is positive; L^b if it is negative.

Rank-Dependent Expected Utility

This is another criterion that is non-linear in probabilities, and is quite popular among the alternatives to expected utility; the authors of our textbook seem to favor it.

Order the consequences $C_1, C_2, \ldots C_n$ in increasing order of preferences. There is a utility function $u(C)$ defined over consequences. There is also a probability weighting function f satisfying $f(0) = 0$, $f(1) = 1$, $f(p) < p$ for all $p \in (0, 1)$. Consider the lottery

$$L = (C_1,\, C_2,\, \ldots\, C_n;\; p_1,\, p_2,\, \ldots\, p_n) \,.$$

This is sure (probability 1) to yield utility at least $u(C_1)$; therefore multiply this by the weight $f(1) = 1$. The increment of utility $[u(C_2) - u(C_1)]$ is received with probability $(1 - p_1)$; therefore multiply it by the weight $f(1 - p_1)$. The next increment $[u(C_3) - u(C_2)]$ is received with probability $(1 - p_1 - p_2)$, and so on. Form the weighted sum

$$f(1)\; u(C_1) + f(1 - p_1)\; [u(C_2) - u(C_1)] + f(1 - p_1 - p_2)\; [u(C_3) - u(C_2)] + \ldots \,.$$

When comparing two lotteries, do this for each, and choose the lottery that yields the higher sum.

To illustrate some of its properties, consider the special case with three outcomes. Write $u_i = u(C_i)$ for brevity. The criterion function is

$$u_1 + f(1 - p_1)\ [u_2 - u_1] + f(p_3)\ [u_3 - u_2]$$

[1] Suppose $u(C) = C$, so an expected utility maximizer with this utility-of-consequences function would be risk-neutral. Consider a rank-dependent-expected-utility maximizer comparing two lotteries where $u_2 = \frac{1}{2}(u_1 + u_3)$ and in lottery L^a we have $p_1^a = p_3^a = 0.5$, $p_2^a = 0$ while in lottery L^b, we have $p_1^b = p_3^b = 0$ and $p_2^b = 1$. The person prefers lottery L^b if

$$u_1 + 1 * [u_2 - u_1] + 0 * [u_3 - u_2] > u_1 + f(\tfrac{1}{2})\ [u_2 - u_1] + f(\tfrac{1}{2})\ [u_3 - u_2]\,,$$

or

$$u_2 - u_1 > f(\tfrac{1}{2})\ [u_3 - u_1]\,.$$

But $u_2 = \frac{1}{2}(u_1 + u_3)$, so $u_2 - u_1 = \frac{1}{2}(u_3 - u_1)$ and the inequality becomes $f(\frac{1}{2}) < \frac{1}{2}$, which is true by assumption. Therefore the assumed shape of $f(p)$ implies risk aversion. Conversely, if we took a different weighting function with $f(p) > p$ for $p \in (0, 1)$, that would yield risk-loving behavior.

[2] To illustrate, suppose further that $f(p) = p^2$, so the criterion function is

$$u_1 + (1 - p_1)^2\ [u_2 - u_1] + (p_3)^2\ [u_3 - u_2]$$

In the probability triangle diagram, the curves where this is constant are ellipses with their center at the point $p_1 = 1$ and $p_3 = 0$, that is, the southeast corner of the triangle. These do not have a uniform fanning out or fanning in property, but they can be consistent with the Allais paradox in special cases.

Induced Preferences

This recognizes that choice among random prospects is not the only thing people do; after they have made the choice, they make some other decisions that affect their eventual outcomes. When we focus solely on choice over random prospects (lotteries), these other choices are behind the scenes. Suppose such other action is labeled z, and the set of feasible actions is Z. That choice will affect the utility of a consequence, so write that utility as $u(C, z)$.

Now consider a lottery

$$L = (C_1,\, C_2,\, \ldots\, C_n;\ p_1,\, p_2,\, \ldots\, p_n)\,.$$

Its expected utility for a given action z is

$$EU(L, z) = \sum_{i=1}^{n}\ p_i\ u(C_i, z)\,.$$

But the decision-maker is allowed to choose this action after choosing the lottery (but before realization of the actual outcome of the uncertainty). Therefore we should see what is the best expected utility he will get from L, after he makes the resulting optimal choice of z. Call this

$$EU^*(L) = \max_{z \in Z} \; EU(L, z) = \max_{z \in Z} \; \sum_{i=1}^{n} \; p_i \; u(C_i, z) \, .$$

Then the choice of the lottery L itself will be governed by $EU^*(L)$. Call this the *induced* utility function, and the preferences over lotteries it represents are *induced preferences.*

What will the indifference curves of this function look like in the probability triangle diagram? The key is to observe that for each z, $EU(L)$ is a linear function of the probabilities. By choosing z optimally, we are taking the upper envelope of all these separate functions corresponding to different z. And the upper envelope of a family of linear functions is a convex function. Therefore $EU^*(L)$ is a convex function of the probabilities. Mathematically, this is the same as the argument in ECO 310 that led to the result that a firm's *dual* or maximized profit function is convex as a function of the output price.

Very roughly speaking, you can think of this subsequent action as something you could do to mitigate the possible bad consequences of the lottery you have just chosen. Then extreme probabilities may not be bad for you.

In the theory of production in ECO 310, production functions are concave, and their contours (isoquants) are concave as seen from the direction of higher output. The situation here is the opposite: the utility function is convex, so its contours (indifference curves) are convex as seen from the direction of increasing utility. Figure 5 illustrates this.

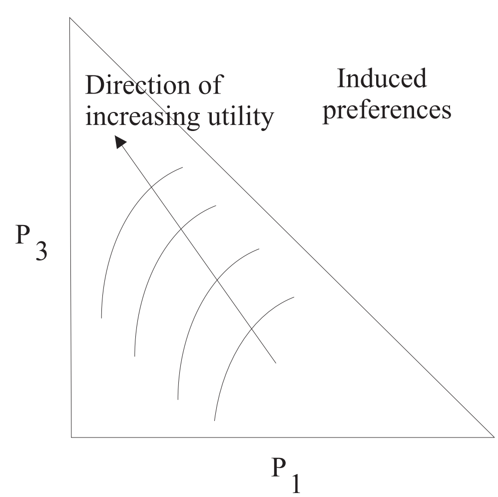

Figure 5: Induced preferences

3. Countercritique - Dutch Books

Many of the counterarguments to critiques of expected utility theory consist of criticisms of the experiments on which the critiques are based, or on interpretations of the findings. But a different countercriticism, developed most prominently by Yaari, is at a conceptual level. It says that if a person's behavior departs from expected utility maximization, then one can present him with a succession of decisions, all of which he takes willingly but which taken together end up costing him money – potentially unlimited amounts of money. If behavior violating expected utility were indeed common, we would see others attempting to profit from it by offering such contracts, called Dutch Books. Since we don't see them, says Yaari, behavior under risk must broadly correspond to expected utility theory.

To get a general idea of the kind of contracts that constitute a Dutch Book, suppose you have non-transitive preferences (with or without the presence of uncertainty; that is not relevant right now). Then there are three or more alternatives, say A, B, C … H, K, such that you prefer B to A, prefer C to B, … prefer K to H, but prefer A to K.[2] I start you out at A. Then I offer to switch you to B, in exchange for a small sum that you are willing to pay for this (less than the money-equivalent of your preference for B over A). That done, I offer to switch you to C in exchange for another small sum. Having proceeded in this way to K, I offer to take you to A in exchange for another small sum; after all you prefer A to K. Then I can start the whole cycle again. You can lose an unbounded amount of money this way, unless you realize that there is something problematic about your preferences and change them so they do become transitive.

In the context of uncertainty, the simplest Dutch Book argument says that if your behavior violates the compound lottery axiom, then I can design a Dutch Book to extract unbounded amounts of expected profit from you. This is simple: If you prefer a compound lottery to a probability-equivalent simple lottery, then I can sell you the compound lottery, in exchange for the simple lottery plus a small sure sum. If you prefer a simple lottery to a probability-equivalent compound lottery, then I can sell you the simple lottery in exchange for the compound lottery and a small sure sum. In either case, the expected value of my trade in lotteries is zero, and I have got a positive sum of money. By playing this game independently often enough I have a large expected gain and you have a large expected loss. Of course the trade is not riskless for me: the net outcome of playing the two lotteries will not be exactly zero every time, unless I can arrange the two lotteries to have perfectly negatively correlated outcomes, so I lose on one precisely when I gain on the other. But I can invite others into this attractive deal, form a syndicate, and spread the risk over all the shareholders with diversified portfolios.

Therefore let us proceed assuming that your behavior conforms to the compound lottery axiom. Next I show that if your behavior violates the independence axiom, then I can design a Dutch Book to make positive expected profit from you.

You violate the independence axiom; suppose you regard L^a and L^b as indifferent, but there is a third lottery L^c and a probability p such that in your preferences,

$$p\ L^c + (1-p)\ L^a \qquad \text{is not indifferent to} \qquad p\ L^c + (1-p)\ L^b\,.$$

[2] Strictly speaking, such preferences are *cyclic* rather than *non-transitive,* but the difference is negligible for our purpose.

To be specific, suppose

$$p\ L^c + (1-p)\ L^a \succ p\ L^c + (1-p)\ L^b\,.$$

Suppose you start at $p\ L^c + (1-p)\ L^b$. I offer to take this from you and give you $p\ L^c + (1-p)\ L^a$ instead, if you pay me a small sum less than the money equivalent of the extent of your preference for the latter lottery, say a dollar just to be definite. You are willing to make this trade. I do this, and then offer a switch to a two-stage lottery, where at the first stage with probability p the prize is L^c and with probability $(1-p)$ it is L^a. Your behavior conforms to the compound lottery axiom, so you are indifferent, and would be willing to make the exchange if I pay you the tiniest sweetener, say a nickel. We make this trade, and I actually run the two-stage lottery. If the outcome is the first one, I give you the L^c. But if it is the latter, I offer to switch you to L^b instead of L^a. This is a matter of indifference to you, so again the tiniest of sweeteners should suffice to get you to accept it.

Let us sum up what has happened. I initially got from you the lottery $p\ L^c + (1-p)\ L^b$. I stand to receive the potential but random winnings of this lottery. Then I played out a first-stage game of tossing a coin with probability of heads p, and paid you L^c or L^b which is again random. The swap and play of lotteries overall has zero *expected* net value to me. Also I received from you a dollar, and gave you two nickels. So my overall expected gain is positive. Once again, it is not without risk: unless I can ensure perfect correlation between my winnings on the lottery $p\ L^c + (1-p)\ L^b$ that I hold, and my obligations to you consisting of the compound lottery with the first stage $p : (1-p)$ yielding L^c or L^b, I cannot ensure that the swap of lotteries will produce zero with certainty. However, once again I can reduce my risk by sharing it with other investors into this venture of getting money from you. And by playing the game repeatedly, I can enjoy unbounded expected gains.

Next comes a more subtle argument: if your indifference curves in the probability triangle diagram have the usual convexity (they look bowed out when viewed from a position of high utility), then I can design a Dutch Book to extract money from you without running any risk, that is, with certainty. This argument comes from an optional item on your reading list: Jerry Green, *Quarterly Journal of Economics,* August 1987.

Figure 6 shows such an indifference curve. Its shape ensures that there exist lotteries

$$L^a = (p_1^a,\, p_2^a,\, p_3^a), \qquad L^b = (p_1^b,\, p_2^b,\, p_3^b)$$

such that $L^a \sim L^b$, but the lottery with probabilities halfway between these two, namely

$$L = (\tfrac{1}{2}\, p_1^a + \tfrac{1}{2}\, p_1^b,\ \tfrac{1}{2}\, p_2^a + \tfrac{1}{2}\, p_2^b,\ \tfrac{1}{2}\, p_3^a + \tfrac{1}{2}\, p_3^b)$$

is preferred to either one of L^a and L^b. Note that L is a simple lottery, but by the compound lottery axiom it can be implemented instead as a 2-stage lottery, call it $L^{(2)}$, where in the first stage a fair coin is tossed to determine whether lottery L^a or L^b is to be played at stage 2.

Furthermore, suppose that for each lottery there exists a sure consequence that you regard as indifferent to the lottery. This requires an infinite number of sure consequences, not just three, so you have to think that we really have an infinite dimensional probability triangle

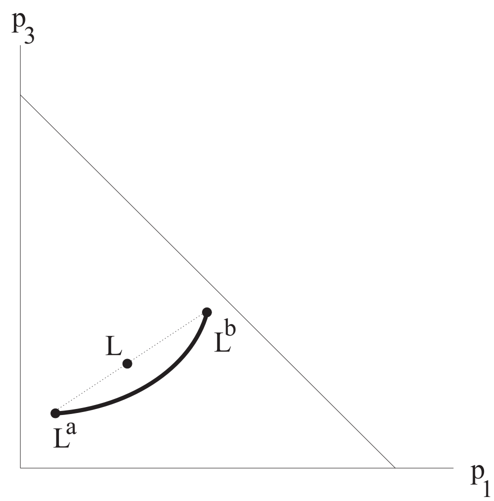

Figure 6: Preferences vulnerable to Dutch Book

and the one shown in the figure is just "schematic." Suppose Q is the sure consequence indifferent to L, and R is the sure consequence that is indifferent to L^a or L^b. So long as you prefer more to less, the sure sum in Q must be bigger than the sure sum in R; say the difference is a dollar just to be definite.

Suppose you are initially at Q. I offer to exchange this for L. You are indifferent, so you should agree; if necessary I can sweeten the deal with a tiny sum, again say a nickel. Now that you have the one-stage lottery L, I offer to exchange it for the two-stage implementation $L^{(2)}$; this is again a matter of indifference so you will agree to it for a nickel sweetener. I run the first stage of $L^{(2)}$, and whatever the outcome, instead of playing out L^a or L^b as the case may be, offer to switch you to R. In each case that is a matter of indifference to you, so the trade costs me another nickel sweetener. The overall outcome is that I have received Q from you, and given you R. So I have gained a dollar and given up three nickels: a positive profit. Note that when I run the first stage, what happens next is independent of the outcome of the first stage, so I am incurring no risk. My profit is non-random.

Note that induced preferences have the opposite convexity, so they are not vulnerable to this Dutch Book.

4. Summing Up

If you find yourself bewildered by the variety of arguments and counterarguments, and by the evidence and rebuttal, that is understandable. Among the active researchers in the area, you will find some are totally convinced of the truth of one of the theories, but different people are committed to different approaches, and many others have not come to any firm conclusion at all.

In our context of choice under uncertainty, expected utility theory has provided the most detailed and richest body of applications so far, but other formulations are gradually catching up at the research level, especially in areas like behavioral finance. In this section of our course I have provided a very brief overview of some of the critiques and alternatives. However, having done so, for the rest of the course we will go back to using expected utility theory in most subsequent topics, because that is what exists at the undergraduate level and also at basic graduate levels of teaching. The textbook in fact postpones discussion of critiques of expected utility to the final chapter, after all the work has already been done using expected utility! I hope some of you will want to explore some of the alternatives in your own independent research.

Reading

Your main readings for this topic should be the Machina (*Journal of Economic Perspectives* 1987) and Kahneman (*American Economic Review* 2003) articles on the reading list. You should also read as much as you can of the Yaari article (xeroxed handout) and the Kahneman-Tversky *Econometrica* 1979 article, but may omit the more technical parts of these. I also highly recommend either of the Camerer items that are optional additional readings.

For the new area of *neuroeconomics,* see the Wikipedia site

`http://en.wikipedia.org/wiki/Neuroeconomics`

and follow whatever leads interest you. This is not needed for our course, but may excite some of you into doing some independent research. One of the leading researchers in this area, Jonathan Cohen, is a professor in our Psychology department, and directs the Center for the Study of Brain, Mind and Behavior and the Program in Neuroscience.