Chapter 11

LIFE AND DEATH CHOICES

1. Introduction

All of the models we have considered to this point, and many we will turn to later, have taken the population, the set of decision makers, the society, as given. The exchange economy model is based on a fixed population of traders. The Pareto criterion says x is superior to y if all the people in a given population like x at least as well as y, and some like it better. The Kaldor criterion is based on a fixed population. Majority voting, which we will study in the next chapter, says x is as good as y if the number of people in a given population who prefer x to y is greater than or equal to the number of people in the same population who prefer y to x. An Arrow social welfare function, which we will analyze in Chapter 13, takes the preferences of each member of a given society and transforms those preferences into a social preference relation.

But what if the population changes? For instance, what if a set of individuals $\{1, 2, \ldots, n\}$ is attempting to choose between alternatives x and y, but x will kill off some of the people, and y will add additional people?

In fact, this is an extremely common question that policy makers and economists face almost every day. For instance: Should a state spend \$5 million replacing a highway if those repairs will likely result in 1 less traffic fatality in the next year? Should a government spend \$10 billion on AIDS drugs if those drugs will prevent 1,000 deaths? Should a government prohibit a sport or leisure activity if that sport creates a 1/6 probability of death per play (e.g., Russian roulette with a 6-chamber revolver)? Should it prohibit a sport or leisure activity if that sport creates a 1/1,000,000 probability of death per day (e.g., downhill skiing)?

Is it better for a country to have a higher population or lower? If it is better to have more people, should this be done by encouraging births, or increasing life expectancy? If it is better to have fewer people, is it better to reduce birth rates or increase deaths?

In this chapter we will look at some partial answers to questions like these. We will start with the standard economic approach to life/death issues, and then we will turn to the more abstract, philosophical approaches.

2. Economic Model – The Money Value of a Life

Placing a money value on a life in legal disputes is an ancient practice. For instance, in the Code of Hammurabi (circa 1750 B.C.) there is a paragraph that reads, in part, "if a citizen has struck a citizen in a quarrel, and has inflicted on him a wound, [and] if he has died as a consequence..., he shall swear, [he struck him unwittingly], ... and ... he shall pay a half-mina of silver." Some lines of *Exodus* suggest money damages for accidental killings in limited circumstances. The Justinian Code provided for money payments to be made in case of accidental killing of slaves. Islamic law distinguishes between intentional and accidental killing, and provides for paying of *diyah* "blood-wit" in the case of accidental killing. Similarly, early custom in pre-Norman Britain put a compensating price (*wirgild*) on deaths.

The modern Anglo-American legal treatment of accidental killing, which started in the mid 19th century, typically provides that dependents of a deceased person may recover for pecuniary losses they suffer, especially lost wages the deceased would have provided. The deceased is primarily viewed as a money making machine. The value of his life is mainly given by lifetime income or earnings, possibly net of expenses needed to maintain the machine (e.g., food, clothing, etc.), possibly discounted to present value, and possibly augmented by the value of non-paid services provided. This can be called the *human capital* approach: the person is valued as a (human) money making machine.

The human capital approach to valuing lives, however, ignores how much the deceased himself would value *being alive*.

How can being alive be valued? Consider the question: "How much is your life worth, in dollars?" What does it mean? It might be a threat: a robber has a gun pointed at you; you have $1,000 in your pocket, and he asks the question. You hand over the $1,000 and he goes away. You have then revealed that you are willing to pay $1,000 (or more) to preserve your life. Or, it might be an opportunity: a benefactor with great wealth approaches you (and your attorney) with an offer: if you sacrifice your life, he will pay \$10 billion to your estate, which will then be distributed as your will provides. You accept the offer, sacrifice your life, and your spouse and children are wealthy as a result. You have now revealed that you are willing to accept \$10,000,000,000 (or less) to sacrifice your life.

Obviously exercises like these will tend to produce wildly disparate numbers. The standard economic measure of the value of a life, in contrast, looks at willingness-to-pay (or willingness-to-accept) for small changes in the probability of death.

The *willingness-to-pay* approach to the value of life works as follows: Suppose one person has an opportunity to reduce his probability of dying by ε, if he participates in some government program, uses some medical procedure, buys some safety enhancement for his car, and so on. Let c be the maximum he would be willing to pay for the given ε. Then we say the willingness-to-pay value of his whole life is c/ε.

Alternatively, a person may be faced with an increase in the probability of dying of ε, because of a riskier job, a hazardous trip, etc. Let c be the minimum he would be willing to accept to compensate him for the given ε. Then we say the willingness-to-accept value of his whole life is c/ε.

(This type of analysis is largely due to Thomas Schelling (1968), E.J. Mishan (1971) and M.W. Jones-Lee (1974). A good survey can be found in W. Kip Viscusi (1993).)

3. A Formal Version of the Economic Model

We will now develop a relatively simple model to show how one individual "computes" the value of his life.

In this model there is just one person, so we will dispense with an identifying subscript. There are two time periods. In period 1, the planning or ex-ante period, he decides on how to allocate his spending. He can spend on consumption, on precaution, or on insurance. Between period 1 and period 2, the ex-post period, events unfold, which leave him either alive, or dead. The probability that he ends up alive in period 2 depends on how much he spends on precaution in period 1. If he is alive, he consumes the amount he chose in period 1. If he is dead, the amount he would have consumed, plus the value of any insurance policy he bought, is bequeathed to his heirs.

We use the following notation:

$$\begin{aligned}
x &= \text{ consumption in period 2 (or part of bequest, if he is dead)} \\
y &= \text{precaution expenditure} \\
z &= \text{insurance expenditure} \\
w = x + y + z &= \text{ initial cash endowment} \\
q(y) &= \text{probability he is alive in period 2} \\
V &= \text{face value of any life insurance policy he buys}
\end{aligned}$$

We assume the $q(y)$ function is nicely behaved: $0 < q(y) < 1$ for all y, $q(y)$ increasing in y, concave, and smooth.

We assume that the cost of life insurance would reflect the actual odds that he will die, so that $z = V \cdot (1 - q(y))$. That is, the price of insurance is "actuarially fair."

We assume our individual reasons as follows: He recognizes there are 2 states of the world for him in period 2, alive, or dead. He has a state-contingent utility function:

$$\begin{cases} f(x) = x^{\alpha} \text{ if alive} \\ g(x+V) = (x+V)^{\alpha} - K \text{ if dead.} \end{cases}$$

Note that α is some fixed parameter, with $0 < \alpha < 1$, and K is some constant.

Let's pause to consider this assumption. The "if alive" part is fairly reasonable, as consumption x increases, he is happier, which makes sense. The particular power function form of the utility function implies that he has diminishing marginal utility from consumption.

The utility "if dead" part requires several comments. First, the $(x + V)^{\alpha}$ part of the function means that our individual contemplates the bequest of what he has not lived to consume (x) plus any life insurance policy (V), just as he would contemplate it if he himself were doing the consuming. That is, we are assuming the same power utility function here. So our person presumably has dependents in mind, or a charity, or some other bequest motive. He cares about consumption by, e.g., his widow and orphans, in much the same way he cares about his own consumption. Second, he (probably) views the dead state as undesirable, with the degree of undesirability captured by the constant K. If K is a large (positive) number, he thinks the dead state is very bad. On the other hand, if it is zero, he views the dead state as similar to the live state, and if K is a negative number, he wants to be a martyr. The position of K in the g function is important. It is outside the power function $(\cdot)^{\alpha}$ and so affects utility in a way that is different than the way money affects utility. This limits the applicability of what is called the complete insurance theorem.

Third, the notion of utility "if dead" is obviously odd: We aren't claiming this is utility *in period 2*. It is not. Once dead, our individual has no utility. However, in period 1, when he is planning, he can think about alternatives (e.g., should I buy a big insurance policy?) contingent on the dead state. The $g(\cdot)$ function is designed to allow rational analysis of such planning.

At this point we combine the utility if alive function $f(x)$ and the utility if dead function $g(x+V)$ into a von Neumann-Morgenstern expected utility function (recall the relevant section of Chapter 1). We assume that in the planning period, our individual maximizes expected utility, which we will call $u(x, y, z)$. Expected utility is the probability of being alive times utility if alive, plus the probability of being dead, times utility if dead. We now have

$$u(x, y, z) = q(y)x^{\alpha} + (1 - q(y))[(x + V)^{\alpha} - K].$$

Substituting for V gives:

$$u(x, y, z) = q(y)x^{\alpha} + (1 - q(y)) \left[\left(x + \frac{z}{1 - q(y)} \right)^{\alpha} - K \right].$$

Our rational planner chooses x, y and z to maximize this function. He is subject to the budget constraint $x + y + z = w$. Also, the variables x and y are constrained to be non-negative, but z need not be. (That is, a person could have a "negative" insurance policy that provides greater consumption if he is alive, but reduces his bequest if he is dead.)

Maximizing this function of 3 variables, subject to the budget constraint, is a somewhat complex exercise. We will not go through all the detailed steps. For interested readers, the procedure is as follows: First, take partial derivatives of $u(x, y, z)$ with respect to the 3 variables. This gives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ and $\frac{\partial u}{\partial z}$. Set these partial derivatives equal to each other, to get "first-order" conditions for the maximum. Use the first-order conditions, and the budget constraint, to characterize the solution. (Note that solving for specific values for x and y would require assuming a specific functional form for $q(y)$ which we will not do.)

As it turns out, given the assumptions listed above, the utility maximizing choices of z and V (i.e., the amount spent on life insurance, and the value of the policy) are both zero. So z and V drop out, and expected utility can therefore be rewritten as

$$u(x, y) = q(y)x^{\alpha} + (1 - q(y))[x^{\alpha} - K].$$

It also turns out that the first-order conditions produce the following equation relating $q'(y)$, the derivative of the $q(y)$ function, x, and y:

$$q'(y) = \frac{\alpha}{K} x^{\alpha-1}.$$

Now recall the modern economic measure of willingness-to-pay value of life (or VOL for short), as described earlier. VOL is the highest c a person is willing to pay, per incremental reduction in his probability of dying ε. That is, VOL $= c/\varepsilon$. But the ratio c/ε has units (change in dollars)/(change in probability). Now note that for our model, the derivative $q'(y)$ has units (change in probability)/(change in dollars). Moreover when our individual is maximizing expected utility subject to a budget constraint, he is increasing y to the point where he is indifferent about that last dollar spent on precaution. That is, he is paying the maximum he would be willing to pay for that last reduction in the probability of dying.

In short, our model produces the modern economic measure of VOL, and it is simply $1/q'(y)$. Therefore, we have

$$\text{VOL } = \frac{K}{\alpha} x^{1-\alpha}.$$

For instance, if $\alpha = 1/2$ (square root utility), if $x = 25,000$ (the right order of magnitude for annual income of a wage-earner in the United States) and $K = 3162$, then VOL $= 1,000,000$ (the right order of magnitude for late 20*th* century empirical studies on VOL in the U.S.)

4. The Broome Paradox

John Broome (1978) attacks the standard economic approach to valuing life in the following way: Suppose the government plans a project that will put some lives at risk. (For instance, a long highway tunnel, or a large bridge, or a military action.) Suppose that, based on past experience, it anticipates 5 deaths from this project. Assuming the formula for VOL developed above, it figures a cost in terms of lives lost of 5 million dollars. Suppose further that, other than the lives lost, the project will cost 20 million, but will produce benefits valued at 50 million.

Cost-benefit analysis then indicates the net project benefit at $50-20-5 = +25$ million, and the project is worthwhile. Suppose the government proceeds with the project, and, as anticipated, 5 workers die. Assume they are Adam, Ben, Charles, Dave and Edward. Consider Adam. He was hired to work on the project, and may have been paid for the additional risk he incurred. However, the additional risk was small, as was his payment to accept it. But we know, ex post, that he did die. We might wonder about the following hypothetical question:

How much cash, say c, would he have required to make him indifferent between the live state and the dead state? Recall that in our model, where individuals are choosing not to buy insurance, the live state utility is x^α and the dead state utility is $x^\alpha - K$. Add c to the dead state x, and set the live state utility equal to the dead state utility. This gives:

$$x^\alpha = (x + c)^\alpha - K.$$

It follows that

$$c = (x^\alpha + K)^{1/\alpha} - x.$$

Using the numbers given above implies

$$c = (25,000^{1/2} + 3162)^2 - 25,000 = 11 \text{ million}.$$

If the cost benefit analysis were redone, knowing that each of the 5 victims would require 11 million to be compensated for his certain death, net project benefit, in millions, would be

$$50 - 20 - 11 \times 5 = -25.$$

So, if we knew which 5 individuals were to die, and if we were to compensate them for the certain losses of their lives, the net benefit of the project would change from +25 million to -25 million.

In short, there is a glaring inconsistency in the valuation of this project, between the ex-ante evaluation (5 statistical lives lost, +25 million net value) and the ex-post valuation (Adam, Ben, Charles, Dave and Edward dead, -25 million net value.)

5. Ex-Ante and Ex-Post

One response to the Broome paradox is to say that it does not matter: Often society must make life or death decisions before knowing *who* will live or die, and in those circumstances it has no choice but to use the ex-ante method. The ex-post valuation is interesting but irrelevant.

Another response is to ask whether or not there exist any approaches to valuing lives that give the same answers, ex-ante and ex-post.

To illustrate, let's consider an example with 100 identical people, all with the utility and VOL functions described above. That is, for utility,

$$\begin{aligned} u(x, y) &= q(y)x^\alpha + (1 - q(y))[x^\alpha - K] \\ &= x^\alpha - (1 - q(y))K. \end{aligned}$$

Assume as before that $x = 25,000$, and $K = 3162$, and so VOL = 1 million. We do not want to have to assume a particular $q(y)$ function.

Instead, let us simply assume that $q(y)$ starts at 0.99. That is, one person is expected to die.

Now assume that the government is deciding between the status quo, which we will call d_1, or, alternatively, a project that will put 15,000 dollars in the pockets of each of the one hundred original people, but will kill off an extra person. We call this alternative d_2.

The money needed to compensate any of our 100 people for this additional 1 percent risk of death is $0.01 \times \text{ VOL } = 10,000$. Therefore the 15,000 more than compensates. That is, ex-ante, d_2 is superior to d_1, based on the dollar amounts. But if we knew *which* additional person was going to die as a result of society's choice of d_2, the necessary compensation for that person would be 11 million, which far exceeds the $15,000 \times 100 = 1.5$ million total project benefit. In short, ex-post d_2 is inferior to d_1.

That is, the money test of d_1 against d_2 is inconsistent. Hello again to the Broome paradox.

But now let's try a similar test with *utilities* rather than *dollars*. In other words, let us become utilitarians, for a change. We will now add together the (identical) von Neumann-Morgenstern utility function of all the people in this society.

Under d_1, ex ante, there are 100 people, all with utility levels

$$.99(25,000)^{1/2} + .01[25,000^{1/2} - 3162] = 126.49.$$

Multiply by 100 gives total utility for society of 12,649. Under d_1, ex-post, there are 99 living people and 1 dead person. The utility levels are

$$(25,000)^{1/2} = 158.1 \text{ for the living,}$$

and

$$(25,000)^{1/2} - 3162 = -3004.2 \text{ for the dead.}$$

Multiplying the first by 99 and the second by 1, and adding all together gives total utility for society of 12,649. That is, using summed utility as the social metric, d_1 is precisely as good ex-ante as ex-post.

Next, consider d_2. Ex-ante, we have 100 people, each with utility level of:

$$.98(25,000 + 15,000)^{1/2} + .02((25,000 + 15,000)^{1/2} - 3162) = 136.76.$$

Multiplying by 100 gives total utility for society of 13,676. Under d_2, ex-post, there are 98 living people and 2 dead. They have utility levels of:

$$(25,000 + 15,000)^{1/2} = 200.0 \text{ for the living, and}$$

$$(25,000 + 15,000)^{1/2} - 3162 = -2962 \text{ for the dead.}$$

Multiplying the first by 98 and the second by 2 gives a total utility for society of 13,676. That is, using summed utility as the metric, d_2 is precisely as good ex-ante as ex-post.

But this means d_1 and d_2 can be unambiguously compared. By the total utility test, d_2 is clearly superior to d_1. No Broome paradox is possible because there is no difference between the ex-ante and ex-post measures.

This example suggests one way to avoid Broome paradox inconsistencies: Use a "utilitarian metric" for measuring the effects of life-death choices, rather than a "money metric."

This possibility is mentioned by Alistair Ulph. The formalization is based on a theorem of Peter Hammond (1981):

Assume there are n people in society, denoted by $i = 1, 2, \ldots, n$. Society can make alternative decisions, $d = 1, \ldots, D$. There are several alternative states of the world, $s = 1, 2, \ldots, S$, whose probabilities depend on the social decision that is taken. We will assume that the decision taken only affects the probabilities of the various states; it does not affect an individual's ex-post utility in a particular state. Let the probability of state s when decision d is taken be $q_s(d)$.

Person i has state-contingent ex-post utility u_{is}. His von Neumann-Morgenstern expected utility function is

$$u_i(d) = \sum_s q_s(d) u_{is}.$$

If society measures welfare using a utilitarian approach, it calculates social welfare as a weighted sum of the $u_i(d)$ terms. (Taking a weighted sum allows counting person i's utility more or less heavily than person j's. If the weights are equal, then $u_i(d)$ and $u_j(d)$ are counted equally.) Let α_i be the weight attached to person i. Let $SW(d)$ represent social welfare. Then, ex ante,

$$SW(d) = \sum_{i=1}^{n} \alpha_i u_i(d) = \sum_{i=1}^{n} \alpha_i \sum_{s=1}^{S} q_s(d) u_{is}.$$

When social welfare is being measured ex-ante, we know which decision d society is taking, but we do not yet know which state of the world will occur.

Ex-post, we do know which state has occured. If state s has occured, ex-post social welfare is

$$\sum_{i=1}^{n} \alpha_i u_{is}.$$

But this state occurred with probability $q_s(d)$. To get an *average* ex-post measure of social welfare, we would have to take a weighted sum over all the states, using the state probabilities as weights. This gives an average ex-post measure:

$$SW(d) = \sum_{s=1}^{S} q_s(d) \sum_{i=1}^{n} \alpha_i u_{is}.$$

Since

$$\sum_{i=1}^{n} \alpha_i \sum_{s=1}^{S} q_s(d) u_{is} = \sum_{s=1}^{S} q_s(d) \sum_{i=1}^{n} \alpha_i u_{is},$$

the ex-ante utilitarian social welfare measure agrees with the ex-post utilitarian social welfare measure. And this is true for *any* ex-post utility functions.

Therefore the cure for Broome-type paradoxes in evaluating life and death choices may be to replace money metrics with utility metrics. But, then again, maybe not, as the next section will show.

6. Problems with Utilitarian Measures of Life/Death Choices

At this point we will drop the complications created by randomness. We assume society can choose among alternative policies $a, b, c, \ldots$. The choice may be made at a particular point in time, at which point there is a population of persons alive, whom we call, as usual, $\{1, 2, \ldots, n\}$. Such a point in time will be called time zero.

Any alternative choice by society, such as a, has implications for the utilities of everyone alive at time zero, and also for the utilities and *the existence* of persons who are not alive at time zero. (For example, the grandchildren and great grandchildren of those who are, at time zero, 10 years old.)

The *timeless population* for a is that set of all people who have been alive, who are alive, and who will be alive, under policy a. We denote that set of people N_a. Obviously it contains $\{1, 2, \ldots, n\}$ as well as many others. The number of people in N_a is called n_a.

Analysis of the choice between alternative a and b can be done in a way which focuses on the time dimension, or in a way that abstracts from it. Our discussion of ex-ante and ex-post, for example, focuses on the time dimension. We will now analyze the choice between alternatives a and b in a way which mainly abstracts from the time dimension. (We do not abstract from time entirely, since we recognize a time zero, and a given population as of that time.) This type of analysis, which deemphasizes the time dimension, is called timeless utilitarianism.

We will now look closely at some standard utilitarian principles, and critique them following the arguments of John Broome (1985).

(1) The total principle. Consider alternative a. For each person in the timeless population for a, measure the utility of his life. Call it $u_i(a)$. Sum these utility numbers for all such people.

This gives

$$\sum_{i \in N_a} u_i(a).$$

Consider an alternative choice b. Do the same. Then compare

$$\sum_{i \in N_a} u_i(a) \text{ and } \sum_{i \in N_b} u_i(b).$$

The total principle holds that the alternative with the higher total is the better alternative.

When writing utility of life numbers like $u_i(a)$, we will follow this standard convention: if a life is worth living, it has a *positive* utility; if it is not worth living, it has a *negative* utility. So $u_i(a) > 0$ means that, under alternative a, person i has a worthwhile life; $u_i(a) < 0$ means that, under alternative a, person i has such an awful life he would be better off dead, and $u_i(a) = 0$ means that, under alternative a, person i is alive, but with a "neutral" life, not so good as to be worth living, not so bad as to be worth dying. Further, if individual i does not *exist* in alternative a, we will use a dash "$-$" to represent his utility, if we need to represent it in an array of utility numbers.

There are two fatal objections to the total principle.

First is the so-called "repugnant conclusion" of Derek Parfit, discussed in his 1984 book *Reasons and Persons.* Consider an alternative a, with a small or moderate sized (timeless) population of persons, all with high levels of utility. Generally there will be an alternative b with a very large (timeless) population of persons with low (but positive) levels of utility, such that b is superior to a. That is, the total principle opts for a huge number of miserable persons (persons with utility barely above zero) over a moderate number of very happy persons.

Second, the total principle attaches the same weights to the person already dead at time 0, and to the person not-yet-born at time 0, as it does to the person alive here on earth. It places no special importance on the *continued* existence of a person alive at time 0. Many people would reject an ethical principal that treats a currently living person the same as a person who is not yet conceived.

(2) The person-restricted principle. This is meant to escape the "repugnant conclusion." We focus on those persons, and only those persons,

who exist under both a and b. That is, let I be the intersection of the sets N_a and N_b. When weighing alternative a against alternative b, compare

$$\sum_{i \in I} u_i(a) \text{ and } \sum_{i \in I} u_i(b).$$

The person-restricted principle opts for the alternative with the higher (person-restricted) total.

This obviously eliminates the possibility that an alternative a with a small number of very happy people will be found inferior to an alternative b with a huge number of miserable people.

However, like the total principle, it attaches the same weight to a currently living person as to a not-yet-born person (who would be born under both a and b).

Moreover, it creates anomalies like the following example, taken from Broome (1985):

Alternative a: An infant is born with severe disabilities. One million dollars might be spent to save his life, but is not spent, and he dies after 3.65 days. The 1 million is spent instead to save the life of a 20-year-old woman, who then lives happily to age 80. The infant's parents conceive another child after its death; the second baby lives happily for 80 years.

Alternative b: An infant is born with severe disabilities. One million dollars is spent to save his life, and he lives happily to age 80. The 20-year old women died at age 20. The infants's parents do not conceive another baby.

We focus on 3 people: the original baby, the replacement baby, and the young woman. We assume that everyone else's utility is the same under a and b. Let's measure utility by simply counting (assumed happy) years of life.

Under the total principle, the comparison is:

	a	b
Original baby	.01	80
Replacement baby	80	–
Woman	80	20
Totals:	160.01	100

So a is a better choice by the total principle.

Under the person-restricted principle, the replacement baby is stricken from the calculation, since he does not exists in both a and b. The result is:

	a	b
Original baby	.01	80
Woman	80	20
Totals:	80.01	100

So b is better by the person-restricted principle. But many people would reject this result. That is, the person-restricted principle seems wrong in this example.

Another objection to the person-restricted principle is this: Again, let I represent the intersection of the sets N_a and N_b. Suppose all the people in I are equally happy with either a or b. Then the person-restricted principle says a and b are equally good. But suppose the people in $N_a \setminus N_b$ (i.e., who exist under a but not under b) are all happy, whereas the people in $N_b \setminus N_a$ (i.e., who exist under b but not under a) are all miserable. Assume the numbers of people in $N_a \setminus N_b$ and $N_b \setminus N_a$ are the same. Then most people would opt for a over b, contrary to the person-restricted principle result.

(3) The average principle. One of the unpalatable things about utilitarianism is its strong bias towards population increase. An obvious way to eliminate that bias is to calculate average utilities, rather than total utilities. Again, consider a and b. Recall that $n_a =$ the number of persons in the timeless population N_a, and $n_b =$ the number of people in N_b.

When weighing alternative a against alternative b, compare

$$\frac{1}{n_a}\sum_{i\in N_a} u_i(a) \text{ and } \frac{1}{n_b}\sum_{i\in N_b} u_i(b).$$

Choose the alternative with the higher average utility level.

Applied to our original baby, replacement baby and woman example, average utilities under the 2 alternatives are:

$$160.01/3 = 53.34 \text{ for } a, \text{ and } 100/2 = 50.00 \text{ for } b.$$

Therefore a is better.

Although this is a logically neat principle, it has flaws. First, it violates a rather natural and appealing axiom called utility independence. (See Charles Blackorby, Walter Bossert and David Donaldson (2003) for a discussion of utility independence and related independence axioms.) *Utility independence* requires that if there are other people in the timeless populations under alternatives a and b, but their utilities are the same in both, then the comparison of a and b should not depend on the utility levels of those other people.

Let's now add one "other person" to modify the example, as follows:

	$\underline{a}$	$\underline{b}$
Original baby	.01	80
Replacement baby	80	$-$
Woman	80	20
"other person"	x	x
Totals:	$160.01 + x$	$100 + x$

Now the average utilities are:

$$\frac{160.01 + x}{4} = 40.00 + \frac{x}{4} \text{ for } a, \text{ and}$$
$$\frac{100 + x}{3} = 33.33 + \frac{x}{3} \text{ for } b.$$

It follows that average utility under a is greater than average utility under b if and only if $80.04 > x$. That is, if the "other person" has utility of, say, 60 in both a and b, a is better: but if he has utility of, say, 100 in both a and b, then b is better! This seems a glaring inconsistency.

But second, it gets even worse : Since these are timeless utilities, the other person may have lived, and died, 2000 years before our babies and woman! (Blackorby et al. call him Euclid, the inventor of geometry, who lived around the $3rd$ century B.C.)

In short, average utilitarianism makes the judgment between a and b depend on utility levels of people who are indifferent between a and b, and those people might have died years ago!

Third, under the average utility principle, it may be an improvement to get rid of relatively unhappy people. Kill the unhappy, or, more consistent with the timeless utilitarian approach, don't let them be born. This is an unpalatable ethical conclusion for some.

(4) The critical level principle. As an alternative to average utilitarianism, consider the following approach, developed in the 1980's by Charles Blackorby and others: To evaluate an alternative a, first figure $u_i(a)$ for all i in N_a, as before. Let $\bar{u}$ be some positive constant. When weighting a against b compare

$$\sum_{i \in N_a} (u_i(a) - \bar{u}) \text{ and } \sum_{i \in N_b} (u_i(b) - \bar{u}).$$

The summation of the terms $(u_i(a) - \bar{u})$ can be viewed intuitively as a calculation of utility *surplus* under a. Critical level utilitarianism says we should choose the alternative with the higher surplus.

Critical level utilitarianism does satisfy utility independence, unlike average utilitarianism.

However, like the average principle it can be criticized because it implies that the unhappy (those with $u_i(a) < \bar{u}$), should be killed, or at least prevented from being born. This is again an unpalatable ethical position for some.

The total principle, the average principle, and the critical level principle can be summarized and compared in the following way:

Let's use TU(a) as shorthand for total utility under a, i.e.,

$$\mathrm{TU}(a) = \sum_{i \in N_a} u_i(a).$$

Then the goal of total principle is to maximize TU(a). The goal of the average principle is to maximize $\frac{1}{n_a}\mathrm{TU}(a)$. The goal of the critical level principle is to maximize $\mathrm{TU}(a) - n_a\bar{u}$. Both the average principle and the critical level principle escape the "repugnant conclusion" of the total principle (of prescribing a huge population of miserably poor people). Average utilitarianism escapes it by looking at the average, rather than the total of the utilities. Critical level utilitarianism escapes it by requiring that a person being added to the population should be viewed as a "good," only if that person's utility exceeds the critical level.

7. The Pareto Principle and Extended Pareto Principles

In judging among alternatives, when the population is fixed and all persons are alive in all alternative states, we can construct a matrix of utilities to help see what is better than what, in terms of the Pareto criterion, the total utility principle, the average utility principle, and so on. We now list in Table 11.1 the people (by number) in the column headings; each row represents a vector of utility levels, one for each person, resulting from a choice by society.

Table 11.1.

	1	2	3	Total	Average
a	1	8	5	14	4.66
b	2	8	5	15	5
c	1	1	1	3	1

Thus b is Pareto superior to a, b is superior to a in terms of total utility and average utility, and c is inferior to a and b in terms of the Pareto criterion, and the utilitarian measures.

We can easily extend this matrix format to accommodate varying populations in a timeless utilitarian framework. We need only to expand the number of columns so as to list every person alive under any alternative. On the line for each alternative, we enter positive utilities for those persons who are alive in that alternative and have lives worth living. Recall that a negative entry would indicate that i is alive but has a life worth avoiding, and a zero would indicate i exists but has a neutral life. A

Table 11.2.

	1	2	3	4	5	...	1000	n_x	Total	Average
a	1	8	5	–	–	...	–	3	14	4.66
b	2	8	5	–	–	...	–	3	15	5.0
c	–	8	5	–	–	...	–	2	13	6.5
d	.02	.02	.02	.02	.02	...	.02	1000	20	.02

dash indicates that person i does not exist under the given alternative. As before, N_x is the population who exist under alternative x, and n_x is the number of persons in N_x. Now our matrix of alternatives may look like Table 11.2 above.

Note that we have changed c and added a d. As before, alternative b is superior to alternative a by Pareto, total utility, average utility, and so on.

Now a comparison of a and d in Table 11.2 shows the intuition of the Derek Parfit's "repugnant conclusion": Total utilitarianism rates d as better than a, even though the people alive in a have very high average utility (4.66) and those alive in d have very low average utility (.02). (Similar comparisons can be made between b and d, and between c and d.)

A comparison of a and c is very useful. Note first that there is not a "transition" from a to c that involves killing person 1. For person 1, alternative c shows a dash. So in c, person 1 has not been killed, rather he was never born.

Now according to some people, there is a list of souls in heaven, and each soul might come to be born on earth, or might not. (If a soul is never born, it would appear in our matrix as a dash.) If a soul is born, it is a good thing, according to some. If an additional person is born, with a positive lifetime utility, and no existing persons *are made worse off*, we will call the change an improvement by the *weak list of souls principle.* If an additional person is born, with a positive lifetime utility and no existing persons are *erased*, we will call the change an improvement by the *strong list of souls principle.*

By the strong list of souls principle, d is best in Table 11.2, and both a and b are superior to c. However, enlightened people generally reject the strong list of souls principle, first, because it leads directly to the "repugnant conclusion," and second, even worse, because it implies that 1000 happy people should be made miserable to allow one miserable person to be born.

The weak list of souls principle is another matter. In a choice between c and a, many would opt for a on the grounds that $c \to a$ is a "Pareto-like" move. Following Blackorby et al. (1984) and others, we will say the following: If a given alternative is modified by adding one person with a positive lifetime utility, while leaving the original population's utility levels unchanged, the change is a *Pareto population improvement.* If the given alternative is modified by adding one person with lifetime utility greater than or equal to the average utility of the original population, while leaving the original population's utility levels unchanged, the change is an *average-utility Pareto population improvement.* If the given alternative is modified by adding one person with lifetime utility greater than or equal to a constant $\bar{u} > 0$, the change is a *critical-level Pareto population improvement.* Note that all these Pareto population improvement criteria are improvements by the weak list of souls principle.

The move from c to a in Table 11.2 is a Pareto population improvement, and a critical-level Pareto population improvement for $\bar{u} \leq 1$. However it fails as an average-utility Pareto population improvement.

The next matrix reveals that both the Pareto population improvement criterion and the critical-level Pareto population improvement criterion are liable to Derek Parfit-like conundrums. For this example, let $\bar{u} = 2$.

Table 11.3.

	1	2	3	4	5	...	1000	n_x	Total	Average
a	1	8	5	–	–	...	–	3	14	4.66
c	–	8	5	–	–	...	–	2	15	6.5
e	.01	8	5	.01	.01	...	.01	1000	22.97	.023
f	2.01	8	5	2.01	2.01	...	2.01	1000	2017	2.02

In Table 11.3, we first compare c and e: By the Pareto population improvement criterion, e is superior to c. But it creates a crowd (997) of miserable people. We conclude from this example that the Pareto population improvement criteria may be unreliable. Now we compare c and f: By the critical-level Pareto population improvement criteria, f is superior to c. The point of this comparison is that this principle will not endorse a crowd of miserable people (people with utility barely above zero), but it will endorse a crowd of people with utilities marginally above the critical level.

8. What's Repugnant About the Repugnant Conclusion?

We should not leave the topic of utilitarian measures over differing populations without a few comments about real-world issues of population size and the resources of the earth.

When the first edition of this book was published, the world population (of humans) was around 4.5 billion. As this second edition is being prepared, it is around 6.5 billion. It is projected to be around 9.0 billion in the year 2050.

Many of the economic resources that are used by humans on earth are obviously physically limited. The land is limited, the seas are limited, fresh water is limited, easily-recoverable oil, natural gas, iron, coal, etc. are limited. It follows that the larger the population, the smaller is the amount of land, sea, fresh water, oil, natural gas, iron, coal, etc. etc., per person. It seems a truism that a great enough population, and small enough per-capita resources, will cause us to suffer someday.

Of course economists have been saying things like these since Thomas Malthus, and we are materially much better off now than we were when, in 1798, he wrote his *Essay on the Principle of Population*. Nonetheless, it is clear that at some level of the human population, human living standards must drop as population rises, because of the finite resources of the planet. (The belief of some that we can export our extra population to other planets or mine other planets for resources, remains a pipe dream.)

But even if we humans can be 9 billion, or 12 billion, or 15 billion strong without threatening our own living standards, there are the non-human living things on earth to consider. And, according to many experts, non-human life has suffered greatly from our growth.

According to biologists such as Edward O. Wilson, humans have precipitated a massive extinction of non-human species, which may become comparable to the great extinction catastrophes of geologic times. We probably hunted to extinction much of the megafauna of North and South America; we hunted to extinction the large flightless birds of New Zealand, we apparently hunted to extinction the megafauna of Madagascar and Australia. We almost hunted the American bison to extinction, and we hunted the passenger pigeon to extinction. Wilson indicates that one fifth of bird species have disappeared in the last 2000 years, because of human action.

But hunting is only one way humans have devastated other species. We have also damaged other species, in isolated environments, particularly islands, by introducing alien species like rats, pigs and snakes. And most important, we have destroyed habitant. When we cut the forest for timber, for farming, or for cattle raising, we often have a massive impact on the plants and animals in those forests, particularly in tropical rain forests. We have already reduced rain forests to around half their prehistoric size, and we are currently cutting them at a rate of 1-2 percent per year. Wilson estimates that at current rates of deforestation, 10 to 25 percent of rain forest species of plants and animals will disappear in 30 years. There is of course a natural background extinction rate for species, but, according to Wilson, "human activity has increased extinction between 1,000 and 10,000 times [the background rate] in the rain forest by reduction in area alone. Clearly we are in the midst of one of the great extinction spasms of geological history." (Wilson, p. 280).

In our discussion of the principles of total utility and average utility above, we noted that the total utility measure might opt for a huge (human) population of miserable people over a small (human) population of happy people, which is the "repugnant conclusion."

For us, one underlying reason such a conclusion is repugnant is that the huge human population may have a huge negative effect on other living things.

In other words, the utility measures we have discussed only account for human utility, and that makes us wary of pushing them too far.

9. Conclusions About Life and Death Choices

In this chapter we have outlined a basic economic value of life model, in which an individual makes decisions about how much to spend to reduce his probability of dying. It is a rather simplistic 2-state uncertainty model, but it does allow some computations of value of life (VOL) numbers.

The economic model raises difficult philosophical questions, however. We have suggested that some of the philosophical objections to the money metric value of life model, particularly the inconsistency between ex-ante and ex-post evaluations, could be met by a utilitarian metric rather than a money metric. But the utilitarian analysis is also full of problems.

As we will see later in this book, there are very serious objections to *almost any* procedure for aggregating the preferences (or ordinal utilities) of a given set of n persons. If it is difficult or impossible to aggregate the preferences of a given population $\{1, 2, \ldots, n\}$, we should probably expect it to be difficult or impossible to aggregate the utilities of arbitrary populations of arbitrary compositions and arbitrary sizes. And so it is.

10. Exercises

1. Suppose an individual spends x on consumption and y on precaution. (There is no life insurance available.) Let the state contingent utility function be

$$\begin{cases} f(x) = x^{\alpha} \text{ if alive} \\ g(x) = x^{\alpha} - K \text{ if dead} \end{cases}$$

   Assume $q(y)$ = probability the individual is alive in period 2.

   Derive an equation for VOL.

2. Suppose the state contingent utility function is

$$\begin{cases} f(x) = \ln x \text{ if alive} \\ g(x) = \ln x - K \text{ if dead} \end{cases}$$

   Assume $q(y)$ = probability the individual is alive in period 2.

   Derive an equation for VOL.

3. Consider the example presented in the discussion of the person-restricted principle:

	a	b
Original baby	.01	80
Replacement baby	80	–
Woman	80	20
Totals:	160.01	100

   Indicate which is better, and why, by the total principle, the average principle, and the critical level principles with $\bar{u} = 20$.

   For what critical level $\bar{u}$ would a and b be equally good?

4. Show that critical level utilitarianism satisfies utility independence.

11. Selected References

(Items marked with an asterisk (*) are mathematically difficult.)

*1. C. Blackorby and D. Donaldson, “Social Criteria for Evaluating Population Change,” *Journal of Public Economics*, V. 25, 1984, pp. 13-33.

As the title suggests, this paper explores ways to evaluate population change. The authors show that a utilitarian social welfare function satisfying certain plausible assumptions must be a (generalized) critical level utility function.

2. C. Blackorby, W. Bossert and D. Donaldson, "The Axiomatic Approach to Population Ethics," *Politics, Philosophy and Economics*, V. 2, 2003, pp. 342-381.

   This is an excellent non-technical survey of the theory of population ethics. The authors lay out axioms of population change, provide clear examples, and review the major principles: classical utilitarianism, critical level utilitarianism, and average utilitarianism. Many variants of these principles are also covered. Blackorby et al. conclude that there are no principles that satisfy all of a set of reasonable axioms, and they discuss the trade offs among the axioms.

   The paper confines mathematical notation to an appendix.

3. J. Broome, "Trying to Value a Life," *Journal of Public Economics*, V. 9, 1978, pp. 91-100.

   Broome lays out his objection to the economic value of life measurement with clarity and elegance.

4. J. Broome, "The Economic Value of Life," *Economica*, V. 52, 1985, pp. 281-294.

   This is another clearly written piece by Broome, in which he describes and criticizes various utilitarian principles for weighing alternatives with life/death implications. It turns out that, according to Broome "none of the principles I considered seems acceptable." Broome opines that the usual practice for valuing life "has no sound basis," and admits that he has no sound alternative to offer in its place. Much of the "Problems with Utilitarian Measures..." section of this chapter comes from this Broome paper.

5. P. Dasgupta, *Human Well-Being and the Natural Environment*, Oxford University Press, 2001.

   This is a deep and complex book on economics and philosophy, touching on human well-being, population principles, sustainable development, and the natural environment. Dasgupta argues that states, particularly poor ones, have mismanaged the natural environment, and that economists have overestimated growth in human well-being by using measures that ignore environmental degradation.

   According to Dasgupta, wealth includes all of manufactured capital, human capital, knowledge, and natural capital, and sustainable development means each generation should bequeath to its successor generation as much capital as it inherited. Dasgupta also revisits classical (total) utilitarianism and average utilitarianism. He categorically rejects the idea of weighing the well-being of a not-yet-conceived person the same as the well-being of a person now alive.

6. P.J. Hammond, "Ex-Ante and Ex-Post Welfare Optimality under Uncertainty," *Economica*, V. 48, 1981, pp. 235-250.

   Hammond proves that if a social welfare function has the property that ex-post efficiency implies ex-ante efficiency, then the function must be a weighted sum of von Neumann-Morgenstern expected utility functions.

7. M.W. Jones-Lee, "The Value of Changes in the Probability of Death or Injury," *Journal of Political Economy*, V. 82, 1974, pp. 835-847.

   Jones-Lee develops a simple and tractable expected utility model for risk of death.

8. E. J. Mishan, "Evaluation of Life and Limb: A Theoretical Approach," *Journal of Political Economy*, V. 79, 1971, pp. 687-705.

   Mishan surveys various methods for valuing loss of life or limb, and opts for compensating variation measures of changes in probabilities of death or injury. These methods should include external effects. Projects should be undertaken if and only if aggregate net benefits, including these measures, are positive.

9. A. Ulph, "The Role of Ex-Ante and Ex-Post Decision, in the Valuation of Life," *Journal of Public Economics*, V. 18, 1982, pp. 265-276.

   Ulph's paper follows in the wave of controversy created by Broome's paper. He notes the ex-ante/ex-post consistency of utilitarian measures of social welfare, but does not advocate utilitarianism. Instead he suggests that cost-benefit analysts "will need to think more carefully about how to capture both the ex-ante and ex-post distributional considerations ... ."

10. W. Kip Viscusi, "The Value of Risk to Life and Health," *Journal of Economic Literature*, V. 31, 1993, pp. 1912-1946.

    A useful survey of many empirical value of life studies. This summarizes labor-market studies on fatal and non-fatal injuries, as well as some non-labor market studies, e.g., on highway speed risks, cigarette smoking risks, and risks from fires. Viscusi also discusses some questionnaire-based studies.

11. E.O. Wilson, *The Diversity of Life*, W.W. Norton & Company, New York, 1992.

A beautifully written book by an eminent biologist on man's impact on non-human life on earth. Wilson has attempted in this and other books to develop an ethical outlook that incorporates what we know about human evolutionary origins. The ethical bottom line is "prudence." We should be aware of our origins as animals, and not destroy the biological world in which we were born.