

*As soon as questions of will or decision or reason or
choice of action arise, human science is at a loss.*

NOAM CHOMSKY

How should we decide? And how *do* we decide? These are the two central questions of Decision Theory: in the *prescriptive (rational)* approach we ask how rational decisions should be made, and in the *descriptive (behavioral)* approach we model the actual decisions made by individuals. Whereas the study of rational decisions is classical, behavioral theories have been introduced only in the late 1970s, and the presentation of some very recent results in this area will be the main topic for us. In later chapters we will see that both approaches can sometimes be used hand in hand, for instance, market anomalies can be explained by a descriptive, behavioral approach, and these anomalies can then be exploited by hedge fund strategies which are based on rational decision criteria.

In this book we focus on the part of Decision Theory which studies choices between alternatives involving risk and uncertainty. *Risk* means here that a decision leads to consequences that are not precisely predictable, but follow a known probability distribution. A classical example would be the decision to buy a lottery ticket. *Uncertainty* or *ambiguity* means that this probability distribution is at least partially unknown to the decision maker.

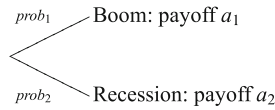
In the following sections we will discuss several decision theories connected to risk. When deciding about risk, rational decision theory is largely synonymous with Expected Utility Theory, the standard theory in economics. The second widely used decision theory is Mean-Variance Theory, whose simplicity allows for manifold applications in finance, but is also a limit to its validity. In recent years, Prospect Theory has gained attention as a descriptive theory that explains actual decisions of persons with high accuracy. At the end of this chapter, we discuss time-preferences and the concept of “time-discounting”.

Before we discuss different approaches to decisions under risk and how they are connected with each other, let us first have a look at their common underlying structure.

2.1 Fundamental Concepts

A common feature of decision theories under risk and uncertainty is that they define so-called *preference relations* between *lotteries*. A lottery is hereby a given set of states together with their respective outcomes and probabilities. A preference relation is a set of rules that states how we make pairwise decisions between lotteries.

Example 2.1 As an example we consider a simplified stock market in which there are only two different states: a boom (state 1) and a recession (state 2). Both states occur with a certain probability $prob_1$ respectively $prob_2 = 1 - prob_1$. An asset will yield a payoff of a_1 in case of a boom and a_2 in case of a recession.



We can describe assets also in the form of a table. Let us assume we want to compare two assets, a stock and a bond, then we have for the payoffs:

State	Probability	Stock	Bond
Boom	$prob_1$	a_1^s	a_1^b
Recession	$prob_2$	a_2^s	a_2^b

The approach summarized in this table is called the “state preference approach”.

If we are faced with a decision between these assets, this decision will obviously depend on the probabilities $prob_1$ and $prob_2$ with which we expect a boom or a recession, and on the corresponding payoffs. However, it might also depend on the *state* in which the corresponding payoff is made. To give a simple example: you might prefer ice cream over a hot cup of tea on a sunny summer day, but in winter this preference is likely to reverse, although the price of ice cream and tea and your budget are all unchanged. In other words, your preference depends directly on the state. It is often a reasonable simplification to assume that preferences over financial goods are *state independent* and we will assume this most of the time. This does not exclude indirect effects: in Example 2.1 a preference might, e.g., depend on the available budget which could be lower in the case of a recession.

In the state independent case, a lottery can be described only by outcomes and their respective probabilities. Let us assume in the above example that $prob_1 = prob_2 = 1/2$. Then we would not distinguish between one asset that yields a payoff of a_1 in a boom and a_2 in a recession and one asset that yields a payoff of a_2 in a boom and a_1 in a recession, since both give a payoff of a_1 with probability $1/2$ and a_2 with probability $1/2$. This is a very simple example for a probability measure on the set of outcomes.¹

To transform the state preference approach into a lottery approach, we simply add the probabilities of all states where our asset has the same payoff. Formally, if there are S states $s = 1, 2, \dots, S$ with probabilities $prob_1, \dots, prob_S$ and payoffs a_1, \dots, a_S , then we obtain the probability p_c for a payoff c by summing $prob_i$ over all i with $a_i = c$. If you like to write this down as a formula, you get

$$p_c = \sum_{\{i=1, \dots, S \mid a_i=c\}} prob_i.$$

To give a formal description of our liking and disliking of the things we can choose from, we introduce the concept of preferences. A *preference* compares lotteries, i.e., probability distributions (or, more precisely, probability measures), denoted by \mathcal{P} , on the set of possible payoffs. If we prefer lottery A over B , we simply write $A \succ B$. If we are indifferent between A and B , we write $A \sim B$. If either of them holds, we can write $A \succeq B$. We always assume $A \sim A$ and thus $A \succeq A$ (reflexivity). However, we should not mix up these preferences with the usual algebraic expressions \geq and $>$: if $A \succeq B$ and $B \succeq A$, this does not imply that $A = B$, which would mean that the lotteries were identical, since of course we can be indifferent when choosing between different things!

Naturally, not every preference makes sense. Therefore in economics one usually considers *preference relations* which are preferences with some additional properties. We will motivate this definition later in detail, for now we just give the definition, in order to clarify what we are talking about.

Definition 2.2 A preference relation \succeq on \mathcal{P} satisfies the following conditions:

- (i) It is complete, i.e., for all lotteries $A, B \in \mathcal{P}$, either $A \succeq B$ or $B \succeq A$ or both.
- (ii) It is transitive, i.e., for all lotteries $A, B, C \in \mathcal{P}$ with $A \succeq B$ and $B \succeq C$ we have $A \succeq C$.

There are more properties one would like to require for “reasonable” preferences. When comparing two lotteries which both give a certain outcome, we would expect

¹We usually allow all real numbers as outcomes. This does not mean that all of these outcomes have to be possible. In particular, we can also handle situations where only finitely many outcomes are possible within this framework. For details see the background information on probability measures in Appendix A.4.

that the lottery with the higher outcome is preferred. – In other words: “More money is better.” This maxim fits particularly well in the context of finance, in the words of Woody Allen:

Money is better than poverty, if only for financial reasons.

Generally, one has to be careful with ad hoc assumptions, since adding too many of them may lead to contradictions. The idea that “more money is better”, however, can be generalized to natural concepts that are very useful when studying decision theories.

A first generalization is the following: if A yields a larger or equal outcome than B in every state, then we prefer A over B . This leads to the definition of *state dominance*. If we go back to the state preference approach and describe A and B by their payoffs a_s^A and a_s^B in the states $s = 1, \dots, S$, we can define state dominance very easily as follows²:

Definition 2.3 (State dominance) If, for all states $s = 1, \dots, S$, we have $a_s^A \geq a_s^B$ and there is at least one state $s \in \{1, \dots, S\}$ with $a_s^A > a_s^B$, then we say that A *state dominates* B . We sometimes write $A \succeq_{SD} B$.

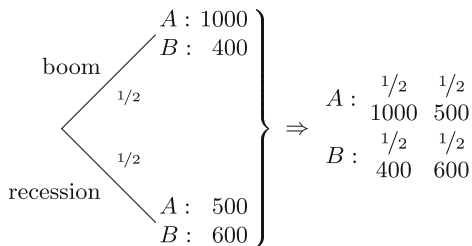
We say that a preference relation \succeq *respects* (or is *compatible with*) *state dominance* if $A \succeq_{SD} B$ implies $A \succeq B$. If \succeq does not respect state dominance, we say that it *violates state dominance*.

In the example of the economy with two states (boom and recession), $A \succeq_{SD} B$ simply means that the payoff of A is larger or equal than the payoff of B in the case of a boom *and* in the case of a recession (in other words always) and at least in one of the two cases strictly bigger.

As a side remark for the interested reader, we briefly discuss the following observation: in the above two state economy with equal probabilities for boom and recession, we could argue that an asset A that yields a payoff of 1000€ in the case of a boom and 500€ in the case of a recession is still better than an asset B that yields 400€ in the case of a boom and 600€ in case of a recession, since the potential advantage of B in the case of a recession is overcompensated by the advantage of A in the case of a boom, and we have assumed that both cases are equally likely (compare Fig. 2.1). However, A does not state-dominate B , since B is better in the recession state. The concept of state-dominance is therefore not sufficient to rule out preferences that prefer B over A . If we want to rule out such preferences, we need to define a more general notion of dominance, e.g., the so-

²It is possible to extend this definition from finite lotteries to general situations: state dominance holds then if the payoff in lottery A is almost nowhere lower than the payoff of lottery B and it is strictly higher with positive probability. See the appendix for the measure theoretic foundations to this statement.

Fig. 2.1 Motivation for stochastic dominance



called *stochastic dominance*.³ We call an asset A *stochastically dominant* over an asset B if for every payoff the probability of A yielding at least this payoff is larger or equal to the probability of B yielding at least this payoff. It is easy to prove that state dominance implies stochastic dominance. We will briefly come back to this definition in Sect. 2.4.

In the following sections we will focus on preferences that can be expressed with a *utility functional*. What is the idea behind this? Handling preference relations is quite an inconvenient thing to do, since computational methods do not help us much: preference relations are not numbers, but – well – relations. For a given set of lotteries, we have to define them in the form of a long list, that becomes infinitely long as soon as we have infinitely many lotteries to consider. Hence we are looking for a method to define preference relations in a neat way: we simply assign a number to each lottery in a way that a lottery with a larger number is preferred over a lottery with a smaller number. In other words: if we have two lotteries and we want to know what is the preference between them, we compute the numbers assigned to them (using some formula that we define beforehand in a clever way) and then choose the one with the larger number. Our analysis is now a lot simpler, since we deduce preferences between lotteries by a simple calculation followed by the comparison of two real numbers. We call the formula that we use in this process a *utility functional*. We summarize this in the following definition:

Definition 2.4 (Utility functional) Let U be a map that assigns a real number to every lottery. We say that U is a *utility functional* for the preference relation \succeq if for every pair of lotteries A and B , we have $U(A) \geq U(B)$ if and only if $A \succeq B$.

In the case of state independent preference relations, we can understand U as a map that assigns a real number to every probability measure on the set of possible outcomes, i.e., $U: \mathcal{P} \rightarrow \mathbb{R}$.

At this point, we need to clarify some vocabulary and answer the question, what is the difference between a *function* and a *functional*. This is very easy: a *function* assigns numbers to numbers; examples are given by $u(x) = x^2$ or $v(x) = \log x$. This is what we know from high school, nothing new here. A *functional*, however, assigns

³Often this concept is called *first order* stochastic dominance, see [Gol04] for more on this subject.

a number to more complicated objects (like measures or functions); examples are the expected value $\mathbb{E}(p)$ that assigns to a probability measure a real number, in other words $\mathbb{E}: \mathcal{P} \rightarrow \mathbb{R}$, or the above utility functional. The distinction between functions and functionals will help us later to be clear about what we mean, i.e. it is important not to mix up utility functions with utility functionals.

Not for all preferences, there is a utility functional. In particular if there are three lotteries A, B, C , where we prefer B over A and C over B , but A over C , there is no utility functional reflecting these preferences, since otherwise $U(A) < U(B) < U(C) < U(A)$. This preference clearly violates the second condition of Definition 2.2, but even if we restrict ourselves to preference relations, we cannot guarantee the existence of a utility function, as the example of a lexicographic ordering shows, see [AB03, p.317]. We will formulate in the next sections some conditions under which we can use utility functionals, and we will see that we can safely assume the existence of a utility functional in most reasonable situations.

2.2 Expected Utility Theory

We will now discuss the most important form of utility, based on the expected utility approach.

2.2.1 Origins of Expected Utility Theory

The concept of probabilities was developed in the seventeenth century by Pierre de Fermat, Blaise Pascal and Christiaan Huygens, among others. This led immediately to the first mathematically formulated theory about the choice between risky alternatives, namely the expected value (or mean value). The expected value of a lottery A having outcomes x_i with probabilities p_i is given by

$$\mathbb{E}(A) = \sum_i x_i p_i.$$

If the possible outcomes form a continuum, we can generalize this by defining

$$\mathbb{E}(A) = \int_{-\infty}^{+\infty} x \, dp,$$

where p is now a probability measure on \mathbb{R} . If, e.g., p follows a normal distribution, this formula leads to

$$\mathbb{E}(A) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

The expected value is the average outcome of a lottery if played iteratively. It seems natural to use this value to decide when faced with a choice between two or more lotteries. In fact, this idea is so natural, that it was the only well-accepted theory for decisions under risk until the middle of the twentieth century. Even nowadays it is still the only one which is typically taught at high school, leaving many a student puzzled about the fact that “mathematics says that buying insurances would be irrational, although we all know it’s a good thing”. (In fact, a person who decides only based on the expected value would not buy an insurance, since insurances have negative expected values due to the simple fact that the insurance company has to cover its costs and usually wants to earn money and hence has to ask for a higher premium than the expected value of the insurance.)

But not only in high schools the idea of the expected value as the sole criterion for rational decision is still astonishingly widespread: when newspapers compare the performance of different pension funds, they usually only report the average return p.a. But what if you have enrolled into a pension fund with the highest average return over the past 100 years, but the average return over your working period was low? More general, what does the average return of the last year tell you about the average return in the next year?

The idea that rational decisions should only be made depending on the expected return was first criticized by Daniel Bernoulli in 1738 [Ber38]. He studied, following an idea of his cousin, Nicolas Bernoulli, a hypothetical lottery A set in a hypothetical casino in St. Petersburg which became therefore known as the “St. Petersburg Paradox”. The lottery can be described as follows: After paying a fixed entrance fee, a fair coin is tossed repeatedly until a “tails” first appears. This ends the game. If the number of times the coin is tossed until this point is k , you win 2^{k-1} ducats (compare Fig. 2.2). The question is now: how much would you be willing to pay as an entrance fee to play this lottery?

If we follow the idea of using the expected value as criterion, we should be willing to pay an entrance fee up to this expected value. We compute the probability p_k that the coin will show “tail” after exactly k times:

$$\begin{aligned} p_k &= P(\text{“head” on 1st toss}) \cdot P(\text{“head” on 2nd toss}) \cdots \\ &\quad \cdots P(\text{“tail” on } k\text{-th toss}) \\ &= \left(\frac{1}{2}\right)^k. \end{aligned}$$

Now we can easily compute the expected return:

$$\mathbb{E}(A) = \sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} 2^{k-1} \left(\frac{1}{2}\right)^k = \sum_{k=1}^{\infty} \frac{1}{2} = +\infty.$$

In other words, following the expected value criterion, you should be willing to pay an arbitrarily large amount of money to take part in the lottery. However, the probability that you win $1024 = 2^{10}$ ducats or more is less than one in a thousand

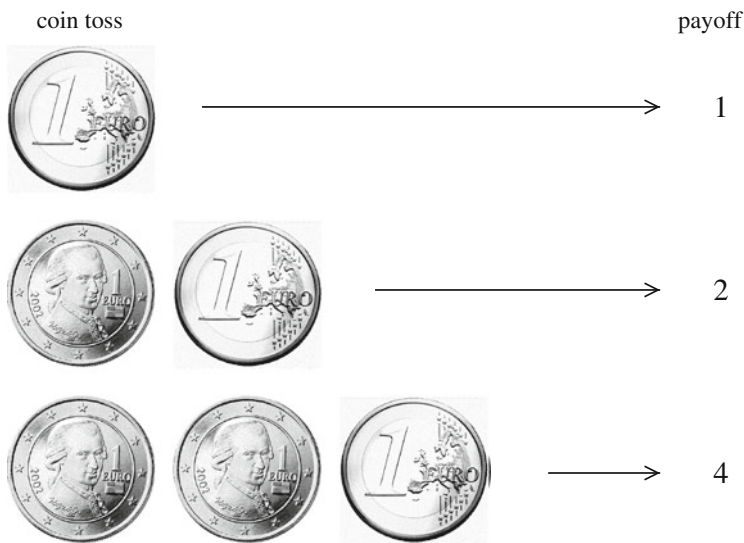


Fig. 2.2 The “St. Petersburg Lottery”

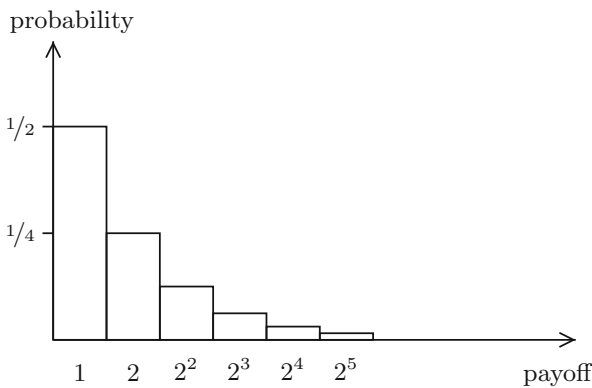


Fig. 2.3 The outcome distribution of the St. Petersburg Lottery

and the infinite expected value only results from the tiny possibility of extremely large outcomes. (See Fig. 2.3 for a sketch of the outcome distribution.) Therefore most people would be willing to pay not more than a couple of ducats to play the lottery. This seemingly paradoxical difference led to the name “St. Petersburg Paradox”.

But is this really so paradoxical? If your car does not drive, this is not paradoxical (although cars are constructed in order to drive), but it needs to be checked, and probably repaired. If you use a model and encounter an application where it produces paradoxical or even plainly wrong results, then this model needs to be checked, and probably repaired. In the case of the St. Petersburg Paradox, the

model was structured to decide according to the expected return. Now, Daniel Bernoulli noticed that this expected return might not be the right guideline for your choice, since it neglects that the same amount of money gained or lost might mean something very different to a person depending on his wealth (and other factors). To put it simple, it is not at all clear why twice the money should always be twice as good: imagine you win one billion dollars. I assume you would be happy. But would you be as happy about then winning another billion dollars? I do not think so. In Bernoulli's own words:

There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount.

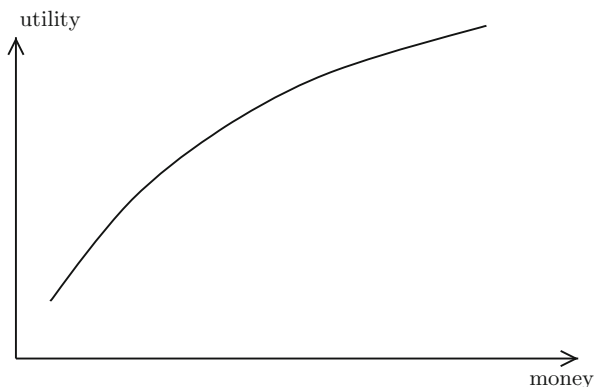
Therefore, it makes no sense to compute the expected value in terms of monetary units. Instead, we have to use units which reflect the usefulness of a given wealth. This concept leads to the *utility theory*, in the words of Bernoulli:

The determination of the value of an item must not be based on the price, but rather on the utility [*"moral value"*] it yields.

In other words, every level of wealth corresponds to a certain numerical value for the person's utility. A utility function u assigns to every wealth level (in monetary units) the corresponding utility, see Fig. 2.4.⁴ What we now want to maximize is the expected value of the utility, in other words, our utility functional becomes

$$U(p) = \mathbb{E}(u) = \sum_i u(x_i)p_i,$$

Fig. 2.4 A utility function



⁴We will see later, how to measure utility functions in laboratory experiments (Sect. 2.2.4), and how it is possible to deduce utility functions from financial market data (Sect. 4.6).

or in the continuum case

$$U(p) = \mathbb{E}(u) = \int_{-\infty}^{+\infty} u(x) dp.$$

Since we will define other decision theories later on, we denote the Expected Utility Theory functional from now on by *EUT*.

Why does this resolve the St. Petersburg Paradox? Let us assume, as Bernoulli did, that the utility function is given by $u(x) := \ln(x)$, then the expected utility of the St. Petersburg lottery is

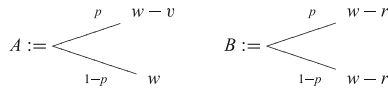
$$\begin{aligned} EUT(\text{Lottery}) &= \sum_k u(x_k)p_k = \sum_k \ln(2^{k-1}) \left(\frac{1}{2}\right)^k \\ &= (\ln 2) \sum_k \frac{k-1}{2^k} < +\infty. \end{aligned}$$

This is caused by the “diminishing marginal utility of money”, i.e., by the fact that $\ln(x)$ grows slower and slower for large x .

What other consequences do we get by changing from the classical decision theory (expected return) to the Expected Utility Theory (EUT)?⁵

Example 2.5 Let us consider a decision about buying a home insurance. There are basically two possible outcomes: either nothing bad happens to our house, in which case our wealth is diminished by the price of the insurance (if we decide to buy one), or disaster strikes, our house is destroyed (by fire, earthquake etc.) and our wealth gets diminished by the value of the house (if we do not buy an insurance) or only by the price of the insurance (if we buy one).

We can formulate this decision problem as a decision between the following two alternative lotteries *A* and *B*, where p is the probability that the house is destroyed, w is our initial wealth, v is the value of the house and r is the price of the insurance:

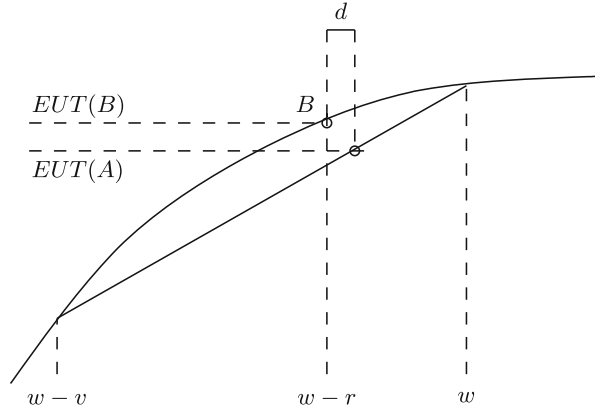


We can also display these lotteries as a table like this:

$$A = \begin{array}{c|cc} \text{Probability} & 1-p & p \\ \hline \text{Final wealth} & w & w-v \end{array}, \quad B = \begin{array}{c|cc} \text{Probability} & 1-p & p \\ \hline \text{Final wealth} & w-r & w-r \end{array}.$$

⁵EUT is sometimes called *Subjective Expected Utility Theory* to stress cases where the probabilities are subjective estimates rather than objective quantities. This is frequently abbreviated by SEU or SEUT.

Fig. 2.5 The insurance problem



A is the case where we do not buy an insurance, in B if we buy one. Since the insurance wants to make money, we can be quite sure that $\mathbb{E}(A) > \mathbb{E}(B)$. The expected return as criterion would therefore suggest not to buy an insurance. Let us compute the expected utility for both lotteries:

$$\begin{aligned} EUT(A) &= (1-p)u(w) + pu(w-v), \\ EUT(B) &= (1-p)u(w-r) + pu(w-r) = u(w-r). \end{aligned}$$

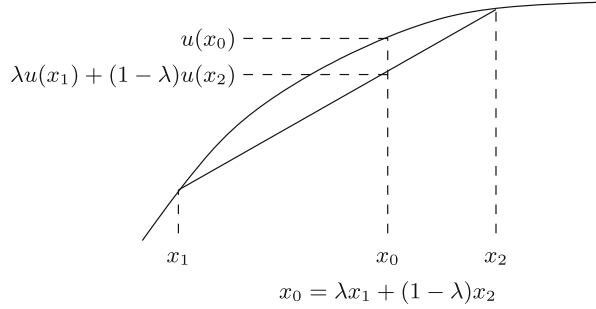
We can now illustrate the utilities of the two lotteries (compare Fig. 2.5) if we notice that $EUT(A)$ can be constructed as the value at $(1-p)v$ of the line connecting the points $(w-v, u(w-v))$ and $(w, u(w))$, since

$$EUT(A) = u(w-v) + (1-p)v \frac{u(w) - u(w-v)}{v}.$$

The expected profit of the insurance d is the difference of price and expected return, hence $d = r - pv$. We can graphically construct and compare the utilities for the two lotteries (see Fig. 2.5). We see in particular, that a strong enough concavity of u makes it advantageous to buy an insurance, but also other factors have an influence on the decision:

- If d is too large, the insurance becomes too expensive and is not bought.
- If w becomes large, the concavity of u decreases and therefore buying the insurance at some point becomes unattractive (assuming that v and d are still the same).
- If the value of the house v is large relative to the wealth, an insurance becomes more attractive.

Fig. 2.6 A strictly concave function



We see that the application of Expected Utility Theory leads to quite realistic results. We also see that a crucial factor for the explanation of the attractiveness of insurances and the solution of the St. Petersburg Paradox is the concavity of the utility function. Roughly spoken, concavity corresponds to risk-averse behavior. We formalize this in the following way:

Definition 2.6 (Concavity) We call a function $u: \mathbb{R} \rightarrow \mathbb{R}$ *concave* on the interval (a, b) (which might be \mathbb{R}) if for all $x_1, x_2 \in (a, b)$ and $\lambda \in (0, 1)$ the following inequality holds:

$$\lambda u(x_1) + (1 - \lambda)u(x_2) \leq u(\lambda x_1 + (1 - \lambda)x_2). \quad (2.1)$$

We call u *strictly concave* if the above inequality is always strict (for $x_1 \neq x_2$).

Definition 2.7 (Risk-averse behavior) We call a person *risk-averse* if he prefers the expected value of every lottery over the lottery itself.⁶

Formula (2.1) looks a little complicated, but follows with a small computation from Fig. 2.6. Analogously, we can define convexity and risk-seeking behavior:

Definition 2.8 (Convexity) We call a function $u: \mathbb{R} \rightarrow \mathbb{R}$ *convex* on the interval (a, b) if for all $x_1, x_2 \in (a, b)$ and $\lambda \in (0, 1)$ the following inequality holds:

$$\lambda u(x_1) + (1 - \lambda)u(x_2) \geq u(\lambda x_1 + (1 - \lambda)x_2). \quad (2.2)$$

We call u *strictly convex* if the above inequality is always strict (for $x_1 \neq x_2$).

⁶Sometimes this property is called “strictly risk-averse”. “Risk-averse” then also allows for indifference between a lottery and its expected value. The same remark applies to risk-seeking behavior, compare Definition 2.9.

Definition 2.9 (Risk-seeking behavior) We call a person *risk-seeking* if he prefers every lottery over its expected value.

We have some simple statements on concavity and its connection to risk aversion.

Proposition 2.10 *The following statements hold:*

- (i) *If u is twice continuously differentiable, then u is strictly concave if and only if $u'' < 0$ and it is strictly convex if and only if $u'' > 0$. If u is (strictly) concave, then $-u$ is (strictly) convex.*
- (ii) *If u is strictly concave, then a person described by the Expected Utility Theory with the utility function u is risk-averse. If u is strictly convex, then a person described by the Expected Utility Theory with the utility function u is risk-seeking.*

To complete the terminology, we mention that a person which has an affine (and hence convex and concave) utility function is called *risk-neutral*, i.e., indifferent between lotteries and their expected return.

As we have already seen, risk aversion is the most common property, but one should not assume that it is necessarily satisfied throughout the range of possible outcomes. We will discuss these questions in more detail in Sect. 2.2.3.

An important property of utility functions is, that they can always be rescaled without changing the underlying preference relations. We recall that

$$U(x_1, \dots, x_S) = \sum_{s=1}^S p_s u(x_s).$$

Then, U is fixed only up to monotone transformations and u only up to positive affine transformations:

Proposition 2.11 *Let $\lambda > 0$ and $c \in \mathbb{R}$. If u is a utility function that corresponds to the preference relation \succeq , i.e., $A \succeq B$ implies $U(A) \geq U(B)$, then $v(x) := \lambda u(x) + c$ is also a utility function corresponding to \succeq .*

For this reason it is possible to fix u at two points, e.g., $u(0) = 0$ and $u(1) = 1$, without changing the preferences. And for the same reason it is not meaningful to compare absolute values of utility functions across individuals, since only their preference relations can be observed, and they define the utility function only up to affine transformations. This is an important point that is worth having in mind when applying Expected Utility Theory to problems where several individuals are involved.

We have learned that Expected Utility Theory was already introduced by Bernoulli in the eighteenth century, but has only been accepted in the middle of the twentieth century. One might wonder, why this took so long, and why this

mathematically simple method has not quickly found fruitful applications. We can only speculate what might have happened: mathematicians at that time felt a certain dismay to the muddy waters of applications: they did not like utility functions whose precise form could not be derived from theoretical considerations. Instead they believed in the unique validity of clear and tidy theories. And the mean value was such a theory.

Whatever the reason, even in 1950 the statistician Feller could still write in an influential textbook [Fel50] on Bernoulli's approach to the St. Petersburg Paradox that he "tried in vain to solve it by the concept of moral expectation." Instead Feller attempted a solution using only the mean value, but could ultimately only show that the *repeated* St. Petersburg Lottery is asymptotically fair (i.e., fair in the limit of infinite repetitions) if the entrance fee is $k \log k$ at the k -th repetition. This implies of course that the entrance fee (although finite) is unbounded and tends to infinity in the limit which seems not to be much less paradoxical than the St. Petersburg Paradox itself. Feller was not alone with his criticism: W. Hirsch writes about the St. Petersburg Paradox in a review on Feller's book:

Various mystifying "explanations" of this paradox had been offered in the past, involving, for example, the concept of moral expectation. . . These explanations are hardly understandable to the modern student of probability.

The discussion in the 1960s even became at times a dispute with slight "patriotic" undertones; for an entertaining reading on this, we refer to [JB03, Chapter 13].

At that time, however, the ideas of von Neumann and Morgenstern (that originated in their book written in 1944 [vNM53]) finally gained popularity and the Expected Utility Theory became widely accepted.

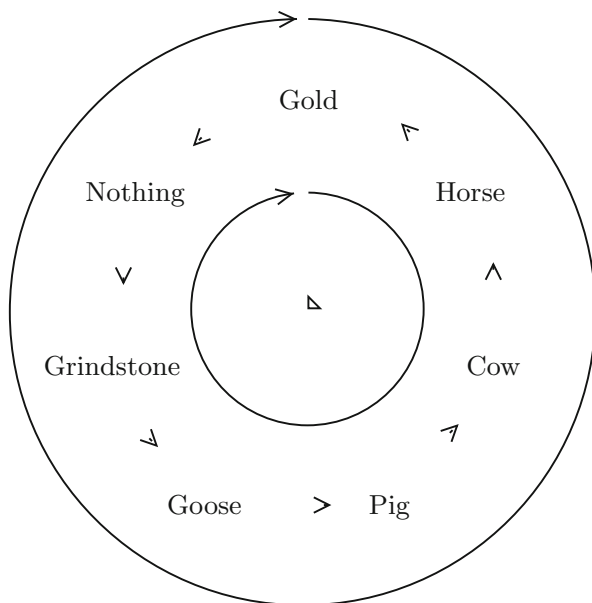
The previous discussions seem to us nowadays more amusing than comprehensible. We will speculate later on some reasons why the time was ripe for the full development of the EUT at that time, but first we will present the key insights of von Neumann and Morgenstern, the axiomatic approach to EUT.

2.2.2 Axiomatic Definition

When we talk about "rational decisions under risk", we usually mean that a person decides according to Expected Utility Theory. Why is there such a strong link between rationality and EUT? However convincing the arguments of Bernoulli are, the main reason is a very different one: we can derive EUT from a set of much simpler assumptions on an individual's decisions. Let us start to compose such a list:

First, we assume that a person should always have *some* opinion when deciding between two alternatives. Whether the person prefers A over B or B over A or whether the person is indecisive, does not matter. But one of these should always be the case. Although this sounds trivial, it might well be that in some context this condition is violated, in particular when moral issues are involved. Generally, and

Fig. 2.7 The cycle of the “Lucky Hans”, violating transitivity



in particular when only financial matters are involved, this condition is indeed very natural. We formulate it as our first *axiom*, i.e., a fundamental assumption on which our later analysis can be based:

Axiom 2.12 (Completeness) *For every pair of possible alternatives, A , B , either $A < B$, $A \sim B$ or $A > B$ holds.*

It is easy to see that EUT satisfies this axiom as long as the utility functional has a finite value.

The next idea is that we should have consistent decisions in the following sense: If we prefer B over A and C over B , then we should prefer C over A . This idea is called “transitivity”. In the fairy tale “Lucky Hans” by the Brothers Grimm, this property is violated, as Lucky Hans happily exchanges a lump of solid gold, that he had earned for 7 years of hard work, for a horse, because the gold is so heavy to carry. Afterwards he exchanges the horse for a cow, the cow for a pig, the pig for a goose, and the goose finally for two knife grinder stones which he then accidentally throws into a well. But he is very happy about this accident, since the stones were so heavy to carry... At the end of the tale he has therefore the same that he had 7 years before – nothing. But nevertheless each exchange seemed to make him happy (Fig. 2.7).

In mathematical terms, “Lucky Hans” preferred B over A , C over B and A over C . Although we might not be blessed with such a cheerful nature, we have to accept that the behavior of some people can be very strange indeed and that the assumption of transitivity might be already too much to describe individuals. However, persons like “Lucky Hans” are probably quite an exception, and the fairy tale would not have

its humorous effect if the audience considered such a transitivity-violating behavior normal. We can therefore feel quite safe by applying this principle, in particular in a prescriptive context.

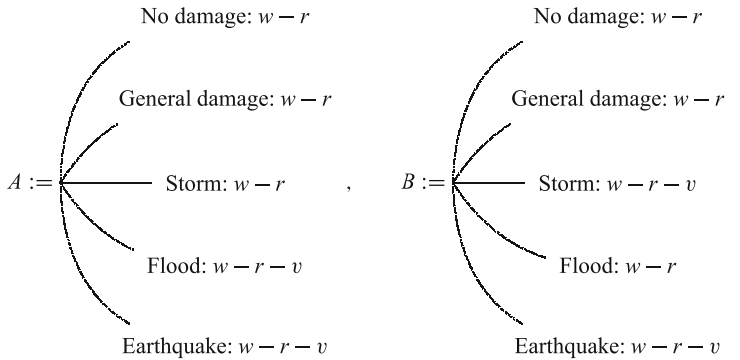
Axiom 2.13 (Transitivity) *For every A, B, C with $A \preceq B$ and $B \preceq C$, we have $A \preceq C$.*

Transitivity is satisfied by EUT and by all other theories that are based on a utility functional, since for these decision theories, transitivity translates into transitivity of real numbers which is always satisfied.

The properties up to now could have been stated for preferences between apples and pears or for whatever one might wish to decide about. It was by no means necessary that the objects under considerations were lotteries. We will now focus on decision under risk, since the following axioms require more detailed properties of the items we wish to compare.

The next axiom is more controversial than the first two. We argue as follows: if we have to choose between two lotteries which are partially identical, then our decision should only depend on the difference between the two lotteries, not on the identical part. We illustrate this with an example:

Example 2.14 Let us assume that we decide about buying a home insurance. There are two insurances on the market that cost the same amount of money and pay out the same amount in case of a damage, but one of them excludes damages by floods and the other one excludes damages by storm. Moreover both insurances exclude damages induced by earthquakes.



If we decide which insurance to buy, we should make our decision without considering the case of an earthquake, since this case (probability and costs) is identical for both alternatives and hence irrelevant for our decision.

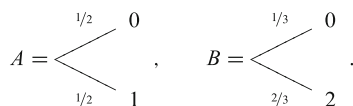
Although the idea to ignore irrelevant alternatives sounds reasonable, it turns out not to be very consistent with experimental findings. We will discuss this when we study descriptive approaches like Prospect Theory in Sect. 2.4. For now, we can

happily live with this assumption, since we are more interested in rational decisions, in other words we follow a prescriptive approach.

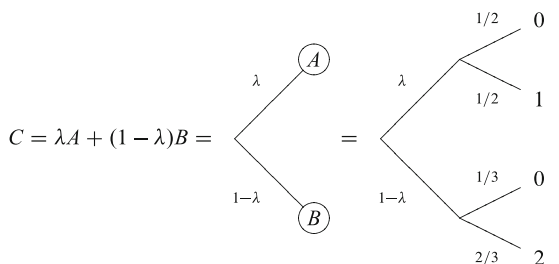
To formulate this axiom mathematically correctly, we need to understand what it means when we combine lotteries.

Definition 2.15 Let A and B be lotteries and $\lambda \in [0, 1]$, then $\lambda A + (1 - \lambda)B$ denotes a new combined lottery where with probability λ the lottery A is played, and with probability $1 - \lambda$ the lottery B is played.⁷

Example 2.16 Let A and B be the following lotteries:



Then the lottery $C := \lambda A + (1 - \lambda)B$ can be calculated as



Alternatively, we can do the same calculation by representing the lottery in a table:

$$A = \begin{array}{c|cc} \text{Probability} & 1/2 & 1/2 \\ \hline \text{Outcome} & 0 & 1 \end{array}, \quad B = \begin{array}{c|cc} \text{Probability} & 1/3 & 2/3 \\ \hline \text{Outcome} & 0 & 2 \end{array}.$$

Then the lottery $C := \lambda A + (1 - \lambda)B$ is

$$C = \lambda A + (1 - \lambda)B = \begin{array}{c|cc} \lambda & 1-\lambda \\ \hline A & B \end{array} = \begin{array}{c|cccc} & \lambda & 1-\lambda & & \\ \hline & \frac{1/2}{A} & \frac{1/2}{B} & \frac{1/3}{A} & \frac{2/3}{B} \\ & 0 & 1 & 0 & 2 \end{array}$$

Both formulations lead to the same result, it is basically a matter of taste whether we write lotteries as tree diagrams or tables. The Independence Axiom allows us now

⁷If the lotteries are given as probability measures, then the notation coincides with the usual algebraic manipulations of probability measures.

to collect compound lotteries into a single lottery, i.e.

$$C \sim \begin{array}{c} \nearrow^{\frac{\lambda}{2} + \frac{1-\lambda}{3}} 0 \\ \nearrow^{\lambda/2} 1 \\ \searrow_{\frac{2(1-\lambda)}{3}} 2 \end{array}$$

or $C \sim \begin{array}{c} \frac{\lambda}{2} + \frac{1-\lambda}{3} \frac{\lambda}{2} \frac{2(1-\lambda)}{3} \\ 0 \quad 1 \quad 2 \end{array}$

A mathematically precise formulation of the Independence Axiom reads as follows:

Axiom 2.17 (Independence) *Let A and B be two lotteries with $A \succ B$, and let $\lambda \in (0, 1]$ then for any lottery C , it must hold*

$$\lambda A + (1 - \lambda)C \succ \lambda B + (1 - \lambda)C.$$

To see that EUT satisfies the Independence Axiom is not so obvious anymore, but the proof is not very difficult. To keep things simple, we assume that the lotteries A, B and C have only finitely many outcomes x_1, \dots, x_n . (A general proof is given in Appendix A.6.) The probability to get the outcome x_i in lottery A is denoted by p_i^A . Analogously, we write p_i^B and p_i^C . We compute

$$\begin{aligned} U(\lambda A + (1 - \lambda)C) &= \sum_{i=1}^n (\lambda p_i^A + (1 - \lambda)p_i^C) u(x_i) \\ &= \lambda \sum_{i=1}^n p_i^A u(x_i) + (1 - \lambda) \sum_{i=1}^n p_i^C u(x_i) \\ &= \lambda U(A) + (1 - \lambda)U(C) \\ &> \lambda U(B) + (1 - \lambda)U(C) \\ &= \lambda \sum_{i=1}^n p_i^B u(x_i) + (1 - \lambda) \sum_{i=1}^n p_i^C u(x_i) \\ &= U(\lambda B + (1 - \lambda)C). \end{aligned}$$

The last axiom we want to present is the so-called “Continuity Axiom”⁸: let us consider three lotteries A, B, C , where we prefer A over B and B over C . Then there should be a way to mix A and C such that we are indifferent between this mix and B . In a precise formulation, valid for finite lotteries⁹:

Axiom 2.18 (Continuity) *Let A, B, C be lotteries with $A \succeq B \succeq C$ then there exists a probability p such that $B \sim pA + (1 - p)C$.*

⁸Sometimes this is also called “Archimedian Axiom”.

⁹In order to make this concept work for non-discrete lotteries, one needs to take a slightly more complicated approach. We give this general definition in Appendix A.6.

One might argue whether this axiom is natural or not, but at least for financial decisions this seems to be a very reasonable assumption. Again, it is not very difficult to see that EUT satisfies the Continuity Axiom. The proof for this is left as an exercise.

Why did we define all these axioms? We have seen that EUT satisfies them (sometimes under little additional conditions like continuity of u), but the reason why they are interesting is a different one: if we don't know anything about a system of preferences, besides that it satisfies these axioms, then they can be described by Expected Utility Theory! This is quite a surprise, since at first glance the definition of EUT as given by Bernoulli seemed to be a very special and concrete concept, but preference relations and the axioms we studied seem to be very general and abstract. Now, both approaches – the direct definition based on economic intuition and the careful, very general approach based only on a small list of natural axioms – lead exactly to the same concept. This was the key insight by Morgenstern and von Neumann [vNM53]. Therefore, utility functions in EUT are often called “von Neumann-Morgenstern utility functions”.

We formulate this central result in the following Theorem that does not follow precisely the original formulation by von Neumann and Morgenstern, but is nowadays the most commonly used version of their result.

Theorem 2.19 (Expected Utility Theory) *A preference relation that satisfies the Completeness Axiom 2.12, the Transitivity Axiom 2.13, the Independence Axiom 2.17 and the Continuity Axiom 2.18, can be represented by an EUT functional. EUT always satisfies these axioms.*

Proof Since the result is so central, we give a sketch of its proof. However, the mathematically inclined reader might want to venture into the realms of Appendix A.6, where the complete proof together with some generalizations (in particular to lotteries with infinite outcomes) is presented.

First, we notice that the (simpler) half of the proof is already done: We have already checked that preference relations which are described by the Expected Utility Theory satisfy all of the listed axioms. What remains is to prove that if these axioms are satisfied, a von Neumann-Morgenstern utility function exists.

Let us consider lotteries with finitely many outcomes x_1, \dots, x_n with $x_1 > x_2 > \dots > x_n$. A sure outcome of x_i can be replaced by a lottery having only the two outcomes x_1 and x_n with some probability q_i and $(1 - q_i)$, as we know from the Continuity Axiom. In other words:

$$x_i \sim \begin{array}{c} q_i \quad x_1 \\ \diagup \quad \diagdown \\ \quad \quad \quad \cdot \\ \diagdown \quad \diagup \\ 1-q_i \quad x_n \end{array}.$$

If we have an arbitrary lottery A with outcomes x_1, \dots, x_n , each of probability p_1^A, \dots, p_n^A , then we can use the Independence Axiom to substitute first the single

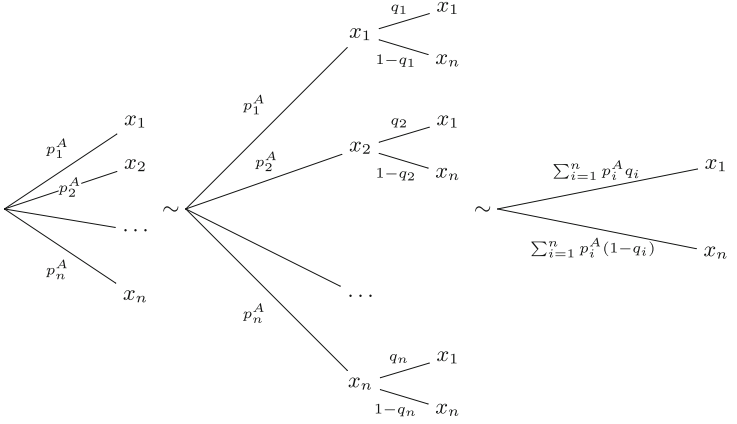


Fig. 2.8 Compound lottery

outcomes by lotteries in x_1 and x_n (using the above equivalence) and then collecting the new lottery into a compound lottery, shown in Fig. 2.8.

If we want to compare two lotteries A and B , we transform them both in this way to get equivalent lotteries A' and B' . Then it becomes very easy for us to decide which lottery is the best: we simply prefer A' over B' if the probability of A' having the better outcome (x_1 or x_n) is larger. To fix ideas, let us assume that x_1 is preferred over x_n , then we just need to compare $U(A) := \sum_{i=1}^n p_i^A q_i$ with $U(B) := \sum_{i=1}^n p_i^B q_i$: if $U(A) > U(B)$, then we prefer A over B ; if $U(B) > U(A)$, then the other way around. Now we can define a utility function u in such a way, that its Expected Utility for any lottery A becomes $U(A)$: simply define $u(x_i) := q_i$, then

$$EUT(A) = \sum_{i=1}^n p_i^A u(x_i) = \sum_{i=1}^n p_i^A q_i. \quad \square$$

Since we convinced ourselves that the listed axioms are all very reasonable, and we tend to say that a *rational* person should obey them, we can conclude that EUT is in fact a good *prescriptive* theory for decisions under risk. However, we have to assume that the utility function considers all relevant effects. Not in all situations the monetary amounts involved are the only relevant effect. Other effects could be based on moral standards, social acceptance etc. EUT as a prescriptive model will work the better the smaller the influence of such factors are that cannot readily be included into the definition of the utility function.

Whether it is also adequate to model behavior of people in real life is an entirely different question, and it will turn out that there are some discrepancies that lead to the development of new *descriptive* theories.

Coming back for a moment to the question, why it took more than 200 years for the development of Expected Utility Theory, a look at other sciences, and in particular mathematics can help us. In fact, the approach by von Neumann and Morgenstern follows a concept that had been used in mathematics intensely at the beginning of the twentieth century and can be summarized as the “axiomatic method”: starting from some fundamental and simple axioms one tries to derive complex theories. Mathematicians stopped accepting objects like the real numbers and merely working with them, but instead developed methods to construct them from simple basic axioms: the natural numbers from some axioms on sets, the rational numbers as fractions of natural numbers, the real numbers as limits of rational numbers and so forth. This was the method that was waiting to be applied to the problems in decision theory under risk. There was also a strong input from psychology which understood at this time that the elementary object of decisions is the *preference* between objects. Von Neumann and Morgenstern (and together with them other scientists who, around the same time, derived similar models) took this as their starting point and used the axiomatic method from mathematics to derive a solid foundation for rational decisions under risk.

We can now even go a step further and say that the results of von Neumann and Morgenstern enable us to avoid any interpretation of the meaning of “utility”. We may not have means to measure a person’s utility, but we do not need to, since it just provides a useful mathematical concept of capturing the person’s preference (which we can observe quite well). We don’t even have to feel bad about using this mathematically convenient framework, since we have proved that it is not so much of an extra assumption, but a natural consequence of reasonable behavior.

To phrase this idea differently: we have at hands two complementary ways of understanding what the Expected Utility Theory is. Summarizing them will help to remember the core ideas of the theory much more than remembering the formula:

- First, we can use Bernoulli’s idea of the utility function that assigns a “real” value to a given amount of money.¹⁰ If we are faced with a decision under risk, we should use the expected value *of this utility* as a natural method to find the more advantageous alternative. This leads to the formula

$$EUT(A) = \mathbb{E}(u(A))$$

for the expected utility of a lottery A .

- Second, we can neglect any potential deep meaning of the utility functions and consider them merely as a convenient and feasible (in realistic situations as defined by the axioms of this section) way of describing the preferences of a rational person.

¹⁰This approach has recently found a revival in the works of Kahneman and others, compare [KDS99].

The precise definition is made in a way that the utility of a lottery A can be computed as convex combination of the utilities of the various outcomes, weighted by their respective probabilities. If these outcomes are x_i and their probabilities are p_i , then this leads to the formula

$$EUT(A) = \sum_{i=1}^n u(x_i)p_i,$$

respectively the generalization to non-discrete probability measures

$$EUT(A) = \int u(x) dp.$$

As we have seen, both approaches lead to the same result.

Looking back on the theory we have derived so far, we are now left with a different, very practical question: we know that we should use EUT with a monotone and continuous utility function u to model rational decisions under risk, but there are plenty of monotone and continuous functions – actually infinitely many. So, which one should we choose? Are there any further axioms that could guide us to select the right one?

2.2.3 Which Utility Functions Are “Suitable”?

We have seen that Expected Utility Theory describes a rational person’s decisions under risk. However, we still have to choose the utility function u in an appropriate way. In this section we will discuss some typical forms of the utility function which have specific properties.

We have already seen that a reasonable utility function should be continuous and monotone increasing, in order to satisfy all axioms introduced in the last section. We have also already discussed that the concavity respectively convexity of the utility function corresponds to risk-averse respectively risk-seeking behavior. It would be nice if one could derive a quantitative measurement for the degree of risk aversion (or risk-seeking) of a person. Since convexity and concavity are characterized by the second derivative of a function (Proposition 2.10), a naive indicator would be this second derivative itself. However, we have seen that utility functions are only characterized up to an affine transformation (Proposition 2.11) which would change the value of u'' . A way to avoid this problem is the standard risk aversion measure, $r(x)$, first introduced by J.W. Pratt [Pra64], which is defined as

$$r(x) := -\frac{u''(x)}{u'(x)}.$$

The larger r , the more a person is risk-averse. Assuming that u is monotone increasing, values of r smaller than zero correspond to risk-seeking behavior, values above zero correspond to risk-averse behavior.

What is the interpretation of r ? The most useful property of r is that it measures how much a person would pay for an insurance against a fair bet. We formulate this as a proposition and give a proof for the mathematical inclined reader:

Proposition 2.20 *Let p be the outcome distribution of a lottery with $\mathbb{E}(p) = 0$, in other words, p is a fair bet. Let w be the wealth level of the person, then, neglecting higher order terms in $r(w)$ and p ,*

$$EUT(w + p) = u\left(w - \frac{1}{2} \text{var}(p)r(w)\right),$$

where $\text{var}(p)$ denotes the variance of p . We could say that the “risk premium”, i.e., the amount the person is willing to pay for an insurance against a fair bet, is proportional to $r(w)$.

Proof We denote the risk premium by a and get $EUT(w + p) = u(w - a)$. Using $EUT(w + p) = \mathbb{E}(u(w + p))$ and a Taylor expansion on both sides, we obtain

$$\begin{aligned} \mathbb{E}(u(w)) + \mathbb{E}(pu'(w)) + \mathbb{E}\left(\frac{1}{2}p^2u''(w)\right) + \mathbb{E}(O(p^3)) \\ = u(w) - au'(w) + O(a^2). \end{aligned}$$

(Here O is the so-called *Landau symbol*, this means that $O(f(x))$ is a term which is asymptotically less or equal to $f(x)$.)

Using $\mathbb{E}(p) = 0$, we get

$$-\frac{1}{2} \text{var}(p)u''(w) = au'(w) - O(\mathbb{E}(p^3)) - O(a^2)$$

and finally $\frac{1}{2} \text{var}(p)r(w) = -a - O(\mathbb{E}(p^3)) + O(a^2)$. □

This result is particularly of interest, since it connects insurance premiums with a risk aversion measure, and the former can easily be measured from real life data.

What values can we expect for r ? Looking at the problems we have studied so far – the St. Petersburg Paradox and insurances – it is natural to assume that risk aversion is the predominating property. However, there are situations in which people behave in a risk-seeking way:

Example 2.21 Lotteries are popular throughout the world. A typical example is the biggest German lottery, the “Lotto” with a turnover of about 25 Million Euro per draw. A lottery ticket of this lottery costs 0.75€ and the chances of winning a major prize (typically in the one million Euro range) are just 0.0000071 %. The chances of

not getting any prize are 98.1 %. Only 50 % of the money spent by the participants is redistributed, the other half goes to the state and to welfare organizations.

Without knowing any more details, it is possible to deduce that a risk-averse or risk-neutral person should not participate in this lottery. Why? To prove our claim, we use the *Jensen inequality*:

Theorem 2.22 (Jensen inequality) *Let $f: [a, b] \rightarrow \mathbb{R}$ be a convex function, let $x_1, \dots, x_n \in [a, b]$ and let $a_1, \dots, a_n \geq 0$ with $a_1 + \dots + a_n = 1$. Then*

$$f\left(\sum_{i=1}^n a_i x_i\right) \leq \sum_{i=1}^n a_i f(x_i).$$

If f is instead concave, the inequality is flipped.

We assume that you have encountered a proof of this inequality before, otherwise you may have a look into a calculus textbook. We refer the advanced reader to Appendix A.4 where we give a general form of Jensen's inequality that allows to generalize our results to non-discrete outcome distributions.

Let us now see, how this inequality can help us prove our statement on lotteries:

We choose as function f the utility function u of a person and assume that u is concave, corresponding to a risk-averse or at least risk-neutral behavior. We denote the lottery with L . The outcomes of L (prizes *plus* the initial wealth of the person *minus* the price of the lottery ticket) are denoted by x_i , their corresponding probabilities by a_i .

Jensen's inequality now tells us that

$$u(\mathbb{E}(L)) = u\left(\sum_{i=1}^n a_i x_i\right) \geq \sum_{i=1}^n a_i u(x_i) = EUT(L).$$

In other words: the utility of the expected return of the lottery is at least as good as the expected utility of the lottery. Now we know that only 50 % of the raised money are redistributed to the participants, in other words, to participate we have to pay twice the expected value of the lottery. Now since $u(2\mathbb{E}(L)) > u(\mathbb{E}(L))$, we conclude that a rational risk-averse or risk-neutral person should not participate in the lottery.

The fact that many people are nevertheless participating is a phenomenon that cannot be too easily explained, in particular since the same persons typically own insurances against various risks (which can only be explained by assuming risk-averse preferences).

A possible explanation is that their utility functions are concave for low values of money, but become convex for larger amounts. This could also explain why other games of chance, like roulette, that allow only for limited prizes, are by far less popular than big lotteries. One could argue that the marginal utility a person

derives from a loss or gain of one Euro is not very high, but by increasing the wealth above a certain threshold, the marginal utility could grow. For instance, by winning one million Euro, a person could be free to stop working or move to a nice and otherwise never affordable house. Although we will see more convincing non-rational explanations of this kind of behavior later, we see that assuming that risk attitudes should follow a standard normalized pattern may not be a very convincing interpretation. We could also think of a more extreme example, taken from a movie:

Example 2.23 In the movie “Run Lola Run”, Mannie, a wanna-be criminal, is supposed to deliver 100,000 Deutsch Marks (50,000€) to his new boss, but loses them on the way. Mannie and his girlfriend Lola have 20 min left to get the money somehow from somewhere, otherwise the boss is going to end Mannie’s career, probably in a fatal way. Unfortunately, they are more or less broke.

The utility function for them will obviously be quite special: above a wealth level of 50,000€ everything is fine (large utility), below that, everything is bad (low utility). It is therefore very likely that their utility function will not be concave. In the movie they are faced with the possibilities of robbing a grocery store, robbing a bank, or gambling in roulette in a casino to earn their money quickly. All three options are obviously very risky and reveal their highly risk-seeking preferences. However, advising them to put the little money they have on a bank account does not seem to be a very rational and helpful suggestion.

We conclude that there are no convincing arguments in favor of a specific risk attitude, other than that risk-averse behavior seems to be reasonable for very large amounts of money, as the St. Petersburg Paradox has taught us. Nevertheless, it is often convenient to do so, and one might argue that “on average” one or the other form could be a reasonable assumption.

One such standard assumption is that the risk aversion measure r is constant for all wealth levels. This is called *Constant Absolute Risk Aversion*, short: CARA. An example for such a CARA utility function is

$$u(x) := -e^{-Ax}.$$

We can verify this by computing $r(x)$ for this function:

$$r(x) = -\frac{u''(x)}{u'(x)} = \frac{A^2 e^{-Ax}}{A e^{-Ax}} = A.$$

Realistic values of A would be in the magnitude of $A \approx 0.0001$.

Since it seems unlikely that risk attitudes are independent of a person’s wealth, another standard approach suggests that $r(x)$ should be proportional to x . In other words, the *relative risk aversion*

$$rr(x) := xr(x) = -x \frac{u''(x)}{u'(x)}$$

is assumed to be constant for all x . We call such function *constant relative risk averse*, short: CRRA. Examples for such functions are

$$u(x) := \frac{x^R}{R}, \quad \text{where } R < 1, R \neq 0,$$

and

$$u(x) := \ln x.$$

Setting $R := 0$ for $\ln x$, we get $rr(x) = 1 - R$ for all of these functions. Typical values for R that have been measured are between -1 and -3 , i.e., an appropriate utility function could be

$$u(x) := -\frac{1}{2}x^{-2}.$$

A subclass of these functions are probably the most widely used utility functions $u(x) := x^\alpha$ with $\alpha \in (0, 1)$. These functions seem to be popular mostly for the sake of mathematical convenience: everybody knows their derivatives and how to integrate them. They are also strictly concave and correspond therefore to risk averse behavior which is often the only condition that one needs for a given application. – In other words, they are the perfect pragmatic solution to define a utility function. But please do not walk away with the idea that these functions are the *only natural* or the *only reasonable* or the *only rational* choice for a utility function! We have seen that things are not as easy and there is in fact no good reason other than convenience to recommend the utility function $u(x) = x^\alpha$.

A generalization of the classes of utility functions introduced so far are utility functions with *hyperbolic absolute risk aversion* (HARA). This class is defined as all functions where the *reciprocal of absolute risk aversion*, $T := 1/r(x)$, is an affine function of x . In other words: u is a HARA function if $T := -u'(x)/u''(x) = a + bx$ for some constants a, b . There is a classification of HARA functions by Merton [MS92]:

Proposition 2.24 *A function $u: \mathbb{R} \rightarrow \mathbb{R}$ is HARA if and only if it is an affine transformation of one of these functions:*

$$v_1(x) := \ln(x + a), \quad v_2(x) := -ae^{-x/a}, \quad v_3(x) := \frac{(a + bx)^{(b-1)/b}}{b - 1},$$

where a and b are arbitrary constants ($b \notin \{0, 1\}$ for v_3). If we define $b := 1$ for v_1 and $b := 0$ for v_2 , we have in all three cases $T = a + bx$.

It is now easy to see that HARA utilities include logarithmic, exponential and power utility functions. (we give an overview in Table 2.1.) Of course, by definition,

Table 2.1 Important classes of utility functions and some of their properties. All belong to the class of HARA functions

Class of utilities	Definition	ARA $r(x)$	RRA $rr(x)$	Special properties
Logarithmic	$\ln(x)$	decr.	const.	“Bernoulli utility”
Power	$\frac{1}{\alpha}x^\alpha, \alpha \neq 0$	decr.	const.	Risk-averse if $\alpha < 1$, bounded if $\alpha < 0$
Quadratic	$x - \alpha x^2, \alpha > 0$	incr.	incr.	Bounded, monotone only up to $x = \frac{1}{2\alpha}$
Exponential	$-e^{-\alpha x}, \alpha > 0$	const.	incr.	Bounded

they contain all CARA and CRRA functions. (v_2 is CARA and v_1 and v_3 for $a = 0$ give all CRRA functions.) To assume that a utility function has to belong to the HARA class is therefore certainly an improvement over more specific ad hoc assumptions, like risk-neutrality. It is, however, only a mathematically convenient simplification. We should not forget this fact, when we use EUT.

Unfortunately, it is not uncommon to read of one or the other class of utility functions as being the only reasonable class. Be careful when encountering such statements! Big minds have erred in such questions: take Bernoulli as an example, who suggested a particular CRRA function (the logarithm) as utility function. He argued that it would be reasonable to assume that the marginal utility of a person is inversely proportional to his wealth level. In modern mathematical terminology $u'(x) \sim 1/x$. Integrating this differential equation, we arrive at the logarithmic function that Bernoulli used to explain the St. Petersburg Paradox. However, is this utility function really so reasonable?

Let us go back to the St. Petersburg Paradox and see whether the solution Bernoulli suggested is really sufficient. Can we make the paradox reappear if we change the lottery? Yes, we can: we just need to change the payoffs to the (even larger) value of e^{2^k} . Then with $u(x) := \ln(x)$ (Bernoulli’s suggestion), we get $u(x_k) = \ln(e^{2^{k-1}}) = 2^{k-1}$ and the same computation as in the case of the original paradox now proves that the expected utility of the new lottery is infinite:

$$EUT = \sum_k u(x_k)p_k = \sum_k 2^{k-1} \left(\frac{1}{2}\right)^k = \sum_k \frac{1}{2} = +\infty.$$

More generally, one can find a lottery that allows for a variant of the St. Petersburg paradox for *every unbounded utility function*, as was first pointed out by Menger [Men34].

There are basically two ways of solving this new paradox, which is sometimes called the “Super St. Petersburg Paradox”. We can understand them, like in the case of the original St. Petersburg Paradox, by comparing the decision theory with a car. If your car does not drive, this might basically be due to two factors: either something is wrong with the car (e.g., no fuel, engine broken...) or something is wrong with the place where you try to drive it (e.g., you are stuck on an icy road).

In the case of a model that could mean that there is either something wrong with the model that needs to be fixed or that you try to apply it at a wrong place, in other words you encountered a restriction to its applicability. In the case of the “Super St. Petersburg Paradox” that leaves us with two ways out:

- We can assume an upper bound on the utility function, take for example $u(x) = 1 - e^{-x}$ which is bounded by 1. In this case, every lottery has an expected utility of less than 1, and therefore there is a finite amount of money that corresponds to this utility value.
- We can try to be a little bit more realistic in the setting of our original paradox, and take into account that a casino would only offer lotteries with a finite expected value, in order to be able to earn money by asking for an entrance fee above this value. Under this restriction, one can prove that the St. Petersburg paradox disappears as long as the utility function is asymptotically concave (i.e., concave above a certain value) [Arr74].

In the second case, we restricted the range of applicable situations (“a car does not drive well on icy roads, so avoid them”). In the first case, we fixed our model to cover even these extreme situations (“always have snow chains with you”).

We formulate this as a theorem:

Theorem 2.25 (St. Petersburg Lottery) *Let p be the outcome distribution of a lottery. Let $u: \mathbb{R} \rightarrow \mathbb{R}$ be a utility function.*

- (i) *If u is bounded, then $EUT(p) := \int u(x) dp < \infty$.*
- (ii) *Assume that $\mathbb{E}(p) < \infty$. If u is asymptotically concave, i.e., there is a $C > 0$ such that u is concave on the interval $[C, +\infty)$, then $EUT(p) < \infty$.*

It is difficult to decide which of the two solutions is more appropriate, an interesting discussion on this can be found in [Aum77]. Considerations in the context of Cumulative Prospect Theory seem to favor a bounded utility function, compare Sect. 2.4.4.

There is another interesting idea that tries to select a certain shape of utility function via an evolutionary approach by Blume and Easley [BE92], see also [Sin02]. There are many experiments for decisions under risk on animals which show that phenomena like risk aversion are much older than humankind. Therefore it makes sense to study their evolutionary development. If the number of offspring of an animal is linearly correlated to the resources it obtained, and if the animal is faced with decisions under risk on these resources, then it can be shown that the only evolutionary stable strategy is to decide by EUT with a *logarithmic utility function*. This is a quite surprising and strong result. In particular, all other possible decision criteria will eventually become marginalized. In this sense EUT with logarithmic utility function would be the only decision model we would expect to observe.

One could also try to apply this idea to financial markets and argue that in the long run all investment strategies that do not follow the EUT maximization

with logarithmic utility function will be marginalized and their market share will be negligible. Hence to model a financial market, we only need to consider EUT maximizer with a logarithmic utility function. – This would certainly be a very interesting insight!

However, there are a couple of problems with this line of argument. First, in the original evolutionary setting, the assumption that the number of offspring is proportional to the resources is a light oversimplification. There is, for instance, certainly a lower bound on the resources below which the animal will simply die and the average number of offspring will therefore be zero, on the other hand, there is some upper bound for the number of offspring. Second, the application to financial markets (as suggested, e.g., in [Len04]) is questionable: under-performance on the stock market does not have to lead to marginalization, since it may be counteracted by adding external resources and the investment time might just not be sufficiently long. New investors will moreover not necessarily implement the same strategies as their predecessors which prevents the market from converging to the theoretically evolutionary stable solution. The idea of using evolutionary concepts in the description of financial markets per se is very interesting, and we will come back to this starting in Sect. 5.7.1, but this concept does not seem to have strong implications for the shape of utility functions.

We have seen that there are plenty of ideas how to choose “suitable” utility functions. We have also found a list of properties (continuous, monotone increasing, either bounded or at least asymptotically concave) that rational utility functions should satisfy. Moreover, we have seen various suggestions for suitable utility functions that are frequently used. However, it is important to understand that there is no single class of functions that can claim to be the “right one”. Therefore the choice of a functional form follows to some extent rather convenience than necessity.

2.2.4 Measuring the Utility Function

When we want to elicit a person’s utility function, we have several possible methods to do so. First, we can rely on real-life data, e.g., from investment or insurance decisions. Second, we can perform laboratory experiments with test subjects. In the latter case, there are various possible procedures, which measure points of the utility function. Using these points, a fit of a function can be made, where usually a specific functional form (for instance x^α) is assumed.

We present here just one of the many methods, the *midpoint certainty equivalent method*. In this method, a subject is asked to state a monetary equivalent to a lottery with two outcomes that each occur with probability $1/2$, compare Fig. 2.9. Such a monetary equivalent (“the price of a lottery”) is called a *Certainty Equivalent (CE)*.

If we set $u(x_0) := 0$ and $u(x_1) := 1$ (which we can do, since u is only determined up to affine transformations), then $u(CE) = 0.5$. We set $x_{0.5} := CE$ and iterate this method by comparing a lottery with the outcomes x_0 and $x_{0.5}$ and probabilities $1/2$ each etc.

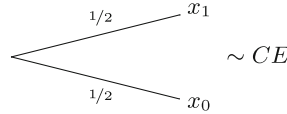


Fig. 2.9 A typical question when measuring a utility function is to ask for a certainty equivalent (CE) for a simple lottery

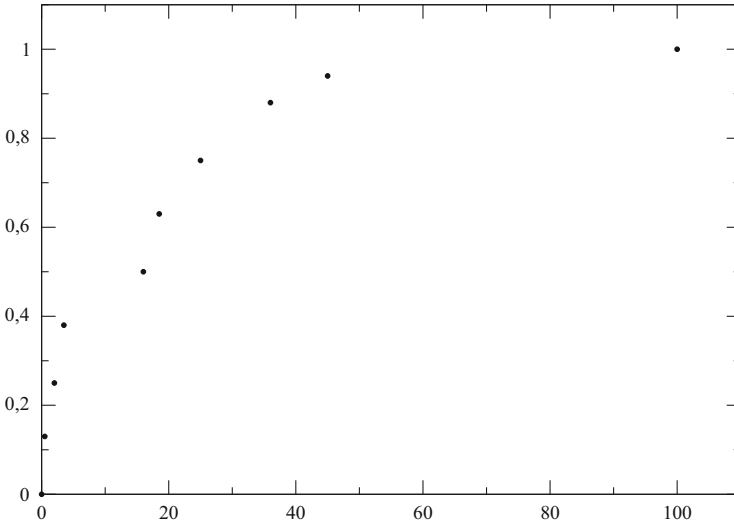


Fig. 2.10 Measured utility function of a test person. x-axis: return of a lottery, y-axis: utility

Let us try this in an example with wealth level w : we set $x_0 := w + 0\text{€}$ and $x_1 := w + 100\text{€}$. The certainty equivalent of a lottery with these outcomes is measured as, say, $w + 15\text{€}$. Thus $x_{0.5} = w + 15$. In the next step we determine the CE of a lottery with outcomes x_0 and $x_{0.5}$. The answer of our test person is 2€ . We then ask for the CE of a lottery with outcomes $x_{0.5}$ and x_1 and get the answer 25€ . Going on with this iteration, we can obtain more data points which ultimately leads to a sketch of the utility function, see Fig. 2.10.

This method has a couple of obvious advantages: it uses simple, transparent lotteries that do not involve complicated, unintuitive probabilities. Moreover, it only needs relatively few questions to elicit a utility function. However, it also has two drawbacks:

- It is not very easy to decide about the certainty equivalent. Pairwise preference decisions are much simpler to do. However, pairwise decisions reveal less information (only yes or no, rather than a numerical value), hence more questions have to be asked in order to get similar results.
- If the test person makes an error, it propagates through the whole experiment, and it is difficult to correct it later on.

There are other methods that avoid these problems, but typically have their own disadvantages. We do not want to discuss them here, but we hope that the example we have given is sufficient to give some ideas on how one can obtain information on this at first glance unascertainable object and what kind of problems this poses.

Assume now that we have measured in an experiment a utility function of a person. The next question we have to ask is, whether EUT is in fact a suitable theory to *describe* these experimental results, since only under this condition our measurements can be used to derive statements about real life situations, e.g., to give advice regarding investment decisions or to model financial markets.

In fact, this question is much more difficult than one might expect. One of the fundamental contributions to this problem has been made by M. Rabin [Rab00] who studied the following question: is it possible to explain the risk aversion that one measures in small stake experiments by means of the concavity of the utility function?

If we have a look on Fig. 2.10, we tend to answer the question affirmatively. The data resembles a function like x^α . However, the x -axis is not the final wealth of the person, but it is just the return of the lotteries, in other words we have to add the wealth w . (In the above example, the person's wealth was roughly 50,000€). Rabin was analyzing such examples a little closer: If we assume a given risk-averse behavior (like rejecting a 50–50 gamble of gaining 105€ or losing 100€) below a certain, not too low wealth level, then it is possible to deduce that very advantageous lotteries would be rejected – regardless of the precise form of the utility function! One can prove, e.g., that if a 50–50 gamble of gaining 105€ or losing 100€ is rejected up to a wealth level of 300,000€, then, at a wealth level of 290,000€, a 50–50 gamble of losing 6000€ and gaining 1.5 Million Euro would still be rejected. This behavior seems to be quite unlikely and not very rational, hence we can conclude that a rational person would not reject the initial offer (lose 100€, gain 105€) up to such a large wealth level.

How does Rabin prove his strong, and somehow surprising result? Without going into the details, we can get an intuition of the result by considering a Taylor expansion of a utility function u at the wealth level w and compute the expected utility of a 50–50 gamble with loss l or gain g :

$$\begin{aligned} & \frac{1}{2}u(w-l) + \frac{1}{2}u(w+g) \\ &= u(w) + \frac{1}{2}(u'(w)(g-l) + \frac{1}{2}u''(w)(g^2-l^2) + O(l^3) - O(g^3)). \end{aligned}$$

Here O is the Landau symbol (see Appendix). Comparing this with $u(w)$, the initial wealth utility, one sees that in order to reject the gamble for all wealth levels or at least for up to a substantial wealth level, $-u''(w)$ has to be sufficiently large. On the other hand $u'(w) > 0$ for all w . This leads to a quickly flattening utility function and to the paradoxical situations observed by Rabin.

The result indicates that EUT might not work well in explaining small stake experiments as illustrated in Fig. 2.10, since it has difficulties in explaining the strong risk aversion that individuals still show – even at relatively large wealth

levels. The simplest way to explain this discrepancy is to use a different “frame”, i.e., to compute the utility function in terms of the potential gains and losses in a given situation, instead of the final wealth. We will see later how this “framing effect” influences decisions and that it is an essential ingredient in modern descriptive theories, in particular in Prospect Theory. It is interesting to observe that this “change of frame” is often intuitively and unintentionally done in textbooks on expected utility theory, a brief search will surely provide the reader with some examples.

Although the paper by Rabin is suggesting to use an alternative approach to describe results of small and medium stake experiments, it has often been misunderstood, in particular in experimental economics, where it is frequently cited as a justification to assume risk-neutrality in experiments. Rabin himself, together with Richard Thaler, admits in a comment [RT02] that

we can see ... how our choice of emphasis could have made our point less clear to some readers

and goes on to remind that risk aversion has been observed in nearly all experiments:

We refer the reader who believes in risk-neutrality to pick up virtually any experimental test of risk attitudes. Dozens of laboratory experiments show that people are averse to far more favorable bets for smaller stakes. The idea that people are not risk neutral in playing for modest stakes is uncontroversial.

He underlines the fact that

because most people are *not* risk neutral over modest stakes, expected utility should be rejected by economists as a *descriptive* theory of decision-making.

Alas, it seems that these clarifications were not heard by everybody.

We will see in Sect. 2.4 what kind of theories are superior as a descriptive model for decisions under risk. Nevertheless it is important to keep in mind that Expected Utility Theory as a *prescriptive* model for *rational* decisions under risk is still largely undisputed. In the next section we will turn our attention to the widely used Mean-Variance Theory which is popular for its “ease of use” that allows fruitful applications where the more complicated EUT is too difficult to apply.

2.3 Mean-Variance Theory

2.3.1 Definition and Fundamental Properties

Mean-Variance Theory was introduced in 1952 by Markowitz [Mar52, Mar91] as a decision criterion for portfolio selection. His key idea was to measure the risk of an asset by only one parameter, the variance σ . Together with the mean μ , these are

the only two parameters that are used in this decision model. Harry Markowitz was awarded the Nobel Prize in 1990 for his pioneering work in financial economics.

In order to make precise what we mean with the “mean-variance approach”, we start with a formal definition:

Definition 2.26 (Mean-Variance approach) A mean-variance utility function u is a utility function $u: \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ which corresponds to a utility functional $U: \mathcal{P} \rightarrow \mathbb{R}$ that only depends on the mean and the variance of a probability measure p , i.e., $U(p) = u(\mathbb{E}(p), \text{var}(p))$.

This definition means that for two lotteries A, B described by the probability measures p_A and p_B , the lottery A is preferred over B if and only if $u(\mathbb{E}(p_A), \text{var}(p_A)) > u(\mathbb{E}(p_B), \text{var}(p_B))$.

The mean is usually denoted by μ , the variance by σ^2 . We can hence express a mean-variance utility functional by writing down the function $u(\mu, \sigma)$.

Of course not every mean-variance utility function is reasonable. – We have already seen in the case of EUT utility functions that for theoretical and practical reasons some properties should be assumed. Most commonly one expects the utility function to be strictly increasing in μ , which corresponds to the “more money is better” maxim. Since σ reflects the risk of a lottery, one usually also assumes that the utility decreases when σ increases. Let us define this precisely:

Definition 2.27 A mean-variance utility function $u: \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is called *monotone* if $u(\mu, \sigma) \geq u(v, \sigma)$ for all μ, v, σ with $\mu > v$. It is called *strictly monotone* if even $u(\mu, \sigma) > u(v, \sigma)$.

We will always assume that u is strictly monotone.

Definition 2.28 A mean-variance utility function $u: \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is called *variance-averse* if $u(\mu, \sigma) \geq u(\mu, \tau)$ for all μ, τ, σ with $\tau > \sigma$. It is called *strictly variance-averse*¹¹ if $u(\mu, \sigma) > u(\mu, \tau)$ for all μ, τ, σ with $\tau > \sigma$.

Often this is assumed as well, but we will not turn this into a general condition. Instead we expect the preference to be risk-averse, i.e., that the expected value of a lottery is always preferred over the lottery itself, compare Definition 2.7. This leads to the following trivial observation:

Remark 2.29 Let u be a mean-variance function. Then the preference induced by u is risk-averse if and only if $u(\mu, \sigma) < u(\mu, 0)$ for all μ, σ . The preference is risk-seeking if and only if $u(\mu, \sigma) > u(\mu, 0)$.

¹¹We use these standard names, although they are not coherent with the use of the similar term “risk-averse” where a *strict* inequality occurs.

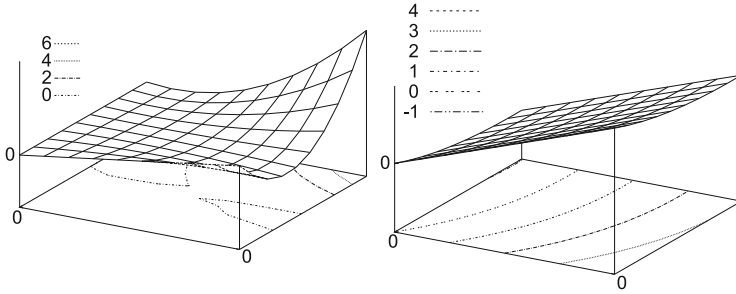


Fig. 2.11 Two mean-variance diagrams

We have found ample evidence for risk-averse behavior in the last section, therefore we consider only mean-variance functions which describe risk-averse behavior.

There is a very convenient way to deduce information on a given mean-variance utility function and the preferences induced by it: the mean-variance diagram, also known as (μ, σ) -diagram. It corresponds to the indifference curves of the utility function on the set of all μ and σ . As an example take the two utility functions¹²

$$u_1(\mu, \sigma) := \mu - \sigma^2, \quad u_2(\mu, \sigma) := 2\mu - 1.3\sigma + 0.5\sigma^2 - 0.054\sigma^3.$$

Their corresponding (μ, σ) -diagrams can be found in Fig. 2.11.

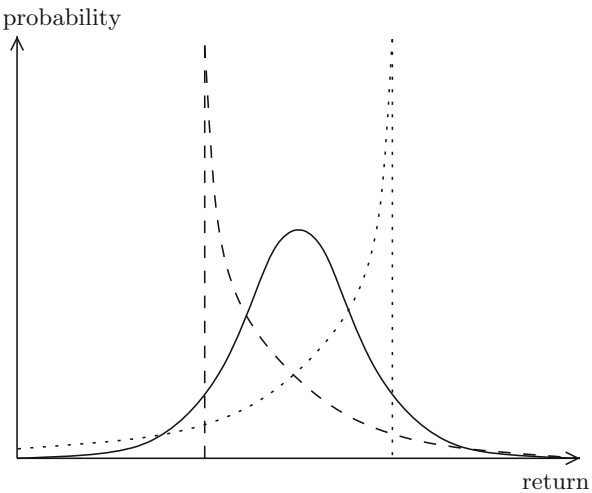
2.3.2 Success and Limitation

The main advantage of the mean-variance approach is its simplicity that reduces the complexity of decisions under risk to only two parameters, the mean μ and the variance σ . This allows us to use (μ, σ) -diagrams in order to characterize the key properties (average return and risk) of an asset. It also allows us to handle complicated models much more easily than with the clumsy EUT. It is therefore no surprise that the Mean-Variance Theory is the most frequently used decision theory by theorists, but also by practitioners in finance, both as a descriptive and prescriptive tool.

On the theoretical side, we will see in the next chapter how this approach can be used to derive the famous Capital Asset Pricing Model which provides easy formulas for price of an asset. We will also see how (μ, σ) -diagrams can be used to derive the Two-Fund-Separation Theorem which states that if everybody is a mean-variance investor and the market is complete and efficient, it is best to hold

¹²The function u_2 is, by the way, an unorthodox suggestion to resolve Allais' Paradox which we will meet in the next section.

Fig. 2.12 Three outcome distributions that are indistinguishable for the Mean-Variance Theory, since their means and variances each agree



a portfolio composed out of a risk-free asset and a representative market portfolio. This outlook highlights the efficiency of the mean-variance approach as a tool in financial economics. However, it also shows its limitation, since practitioners are obviously not following this result, and we may assume that they have reasons.

On the practical side, we mention that banks are usually providing clients with two main informations on assets: the average return and the risk, the latter usually measured as variance.

Although the practical use of an easy method to solve complex problems is surely valuable, there are nevertheless certain problems and limitations of the Mean-Variance Theory. Practitioners sometimes raise the question whether the variance is really an appropriate tool to measure risk. As a simple – albeit more academic – example take, e.g., the following two assets which have identical mean and variance and are hence considered to be equal by the mean-variance criterion:

$A :=$	payoff	0€	1010€
	probability	99.505 %	0.495 %
$B :=$	payoff	-1000€	10€
	probability	0.495 %	99.505 %

There are obviously important reasons why one would like to prefer either A or B , but it seems worthwhile to distinguish both assets! This also holds true for more realistic distributions, compare Fig. 2.12.

A practical application is that banks who are only reducing their risk as measured by the variance might still accept a *very small* probability of an *enormous* loss that might even lead to a bankruptcy. A modern risk management, however, would rather accept a *substantial* risk of a *small* loss which can still be mitigated, even if the associated variance is larger. Another typical practical problem occurs with options which tend to have very skewed payoff distributions. Here the variance as

risk measure does not distinguish the upside risk of making a large profit with low probability from the downside risk of losing a lot with low probability.

There are several other methods to measure risk, like *value at risk* which measures the value of which the payoff falls short in only $n\%$ of the cases or the *expected tail loss* which measures the expected loss that occurs in the worst $n\%$ of the cases. All these practical modifications have in common that they aim to measure the risk with a single quantity, but that they replace the variance by a more sophisticated measure.

Besides these practical problems there are also strong theoretical limitations of the mean-variance approach. The strongest is the so-called “Mean-Variance Paradox”. We formulate it as a theorem:

Theorem 2.30 (Mean-Variance Paradox) *For every continuous mean-variance utility function $u(\mu, \sigma)$ which corresponds to a risk-averse preference, there exist two assets A and B where A state dominates B , but B is preferred over A .*

Proof Let us construct an explicit example, where for simplicity we assume that u is strictly monotone. Consider for $N \geq 1$ the following lottery

$$A_N := \begin{array}{c} \text{payoff in } \mathbb{E} \quad 0 \quad N \\ \text{probability} \quad 1 - \frac{1}{N^2} \quad \frac{1}{N^2} \end{array}.$$

The expected value of A_N is

$$\mathbb{E}(A_N) = \left(1 - \frac{1}{N^2}\right) \cdot 0 + \frac{1}{N^2}N = \frac{1}{N}.$$

The variance can now be easily computed as

$$\begin{aligned} \text{var}(A_N) &= \left(1 - \frac{1}{N^2}\right) \frac{1}{N^2} + \frac{1}{N^2} \left(N - \frac{1}{N}\right)^2 \\ &= \frac{1}{N^2} - \frac{1}{N^4} + 1 - \frac{2}{N^2} + \frac{1}{N^4} = 1 - \frac{1}{N^2}. \end{aligned}$$

Now we compare this with the mean and variance of the lottery A_0 that always gives a payoff of zero:

$$\mathbb{E}(A_0) = 0, \quad \text{var}(A_0) = 0.$$

If N becomes large, its mean value converges to zero, whereas its variance converges to 1. Since u is continuous and risk-averse, this implies that

$$U(A_N) = u\left(\frac{1}{N}, 1 - \frac{1}{N^2}\right) \rightarrow u(0, 1) < u(0, 0) = u(A_0).$$

Therefore, we can choose N large enough, such that the inequality

$$u\left(\frac{1}{N}, 1 - \frac{1}{N^2}\right) < u(0, 0)$$

holds. Since A_N gives a payoff of zero in states with the total probability $1 - \frac{1}{N^2}$, but a positive payoff N in states with a total probability $\frac{1}{N^2}$ which is strictly larger than zero, but A_0 gives in both cases a zero payoff, A_N state dominates A_0 . However, we have just proved that *any* mean-variance utility (which satisfies the initial assumptions) would prefer A_0 over A_N . Setting $A := A_N$ and $B := A_0$ we have proved the theorem. \square

Let us stop here for a moment and think about what we have proved right now: there are two assets, one which never gives any profit, and the other one which does, although with a small probability, and besides that never loses you any money. Sure, you would prefer the latter one! After all, it poses no risk for you. But this is wrong, if you define “risk” as variance. We learn from this that the variance is *really* not a particularly good measure for risk.

Another interesting fact about Mean-Variance Theory which follows directly from the Mean-Variance Paradox is that it does not satisfy the Independence Axiom (compare Definition 2.17):

Corollary 2.31 *Every strictly monotone and risk-averse Mean-Variance Utility violates the Independence Axiom.*

Proof Take the lottery A_N as constructed in the last proof, such that A_0 is preferred over A_N . Both lotteries have a common part: with a probability of $1 - 1/N^2$ they both yield an outcome of zero. Only in the remaining cases (with probability $1/N^2$) they differ: whereas A_N gives an outcome of N , A_0 still gives only 0. If the Independence Axiom were satisfied, we could neglect the common part, and the preference relation would carry over to the remaining cases. However, these lotteries correspond to a sure gain of N or a sure gain of zero, and according to strict monotonicity the gain of N would be preferred. \square

We remark that both assumptions (strict monotonicity and risk-averseness) are indeed necessary requirements for the corollary, since they exclude the special cases of risk-neutral EUT and indifference to mean.

It is also possible to illustrate the violation of the Independence Axiom on a simple example, sometimes referred to as “common ratio effect”:

Example 2.32 Consider four investment alternatives A, B, C and D that yield returns of 2%, 4% or 6% with the probabilities given in Table 2.2. We can list mean and variance of the four investments and then compute their Mean-Variance utility which we choose for simplicity as $U(\mu, \sigma) := \mu - \sigma^2$. In this way we obtain Table 2.2.

Table 2.2 Payoff probabilities for the four hypothetical investments A, B, C and D

Investment	2 %	4 %	6 %
A	0.2	0	0.8
B	0	1	0
C	0.8	0	0.2
D	0.75	0.25	0

Table 2.3 Mean μ , variance σ^2 and $U(\mu, \sigma) = \mu - \sigma^2$ for the four investments from Table 2.2

Investment	Mean μ	Variance σ^2	$\mu - \sigma^2$
A	5.2	2.56	2.64
B	4	0	4
C	2.8	2.56	0.24
D	2.5	2.75	0

This implies that Mean-Variance Theory with the utility $U(\mu, \sigma^2) = \mu - \sigma^2$ predicts the preference pattern $B \succ A \succ C \succ D$ (Table 2.3).

Let us take a closer look at the investments. Then we see that the lottery C is equivalent to playing lottery A with a probability of 1/4 and getting 2 % with a probability of 3/4. Similarly, lottery D is equivalent to playing lottery B with a probability of 1/4 and getting 2 % with a probability of 3/4. Thus, the Independence Axiom would imply that if C is preferred over A, then B had to be preferred over C. The Mean-Variance utility from above, however, shows a different pattern of preferences, thus the Independence Axiom is violated.

That the pattern of preferences predicted by Mean-Variance Theory contradicts Expected Utility Theory, can also be seen directly by a short computation: denote $x := u(2\%)$, $y := u(4\%)$, $z := u(6\%)$. Then $B \succ A$ implies $0.2x + 0.8z < y$ and $C \succ D$ implies $0.8x + 0.2z > 0.75x + 0.25y$ or $0.05x + 0.2z > 0.25y$. Multiplying the last inequality by four gives a contradiction, thus the preference pattern cannot be explained by Expected Utility Theory.

In Sect. 2.5 we compare EUT and Mean-Variance Theory and we will see that there are in fact certain cases, where the problems we have encountered cannot occur and Mean-Variance Theory even becomes a special instance of EUT. In general, however, we need to keep in mind that there is always a risk to apply the mean-variance approach to general situations: beware of being too credulous when applying Mean-Variance Theory!

2.4 Prospect Theory

So, how do people *really* decide? As if they were maximizing their expected utility? Or as if they were following the mean-variance approach? Or do they deviate from both models and decide in a random manner that makes it completely impossible to predict their decisions beforehand? – It turns out that none of these is the case. In this section we will present models that describe actual decisions quite well.

2.4.1 Origins of Behavioral Decision Theory

Although the axioms of Expected Utility Theory were so convincing that we refer to a behavior described by this model as “rational”, it is nevertheless possible to observe people deviating systematically from this rational behavior. One of the most striking examples is the following (often called “Asian disease”):

Example 2.33 Imagine that your country is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. Which of the two programs would you choose?

The majority (72 %) of a representative sample of physicians preferred program A, the “safe” strategy. Now, consider the following, slightly different problem:

Example 2.34 In the same situation as in Example 2.33 there are now instead of A and B two different programs C and D: If program C is adopted, 400 people will die. If program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. Which of the two programs would you favor?

In this case, the large majority (78 %) of an equivalent sample preferred the program D. – Obviously, it would be cruel to abandon the lives of 400 people by choosing program C!

You might have noticed already that both decision problems are exactly identical in contents. The only difference between them is how they are formulated, or more precisely how they are *framed*. Applying EUT cannot explain this observation, neither can Mean-Variance Theory. Moreover, it would not help to modify our notion of a rational decider to capture this “framing effect”, since any rational person should definitely *not* make a difference between the two identical situations. Let us have a look on another classical example of a deviation from rational behavior.¹³

Example 2.35 In the so-called “Allais paradox” we consider four lotteries (A, B, C and D). In each lottery a random number is drawn from the set $\{1, 2, \dots, 100\}$ where each number occurs with the same probability of 1 %. The lotteries assign outcomes to every of these 100 possible numbers (states), according to Table 2.4.

¹³This example might remind the reader of Example 2.32 that demonstrated how Mean-Variance Theory can lead to violations of the Independence Axiom.

Table 2.4 The four lotteries of Allais' Paradox

Lottery A	State	1–33	34–99	100
	Outcome	2500	2400	0
Lottery B	State	1–100		
	Outcome	2400		
Lottery C	State	1–33	34–100	
	Outcome	2500	0	
Lottery D	State	1–33	34–99	100
	Outcome	2400	0	2400

The test persons are asked to decide between the two lotteries A and B and then between C and D. Most people prefer B over A and C over D.

This behavior is not rational, although this time it might be less obvious. The axiom that most people violate in this case is the Independence Axiom. We can see this by neglecting in both decisions the states 34–99, since they give the same result each. What is left (the states 1–33 and the state 100) are the same for both decision problems. In other words, the part of our decisions which is independent of irrelevant alternatives, is the same when deciding between A and B and when deciding between C and D. Hence, if we prefer B over A we should also prefer D over C, and if we prefer C over D, we should also prefer A over B.

We have already encountered other observed facts that can be explained with EUT only under quite delicate and even painstaking assumptions on the utility function:

- People tend to buy insurances (risk-averse behavior) and take part in lotteries (risk-seeking behavior) at the same time.
- People are usually risk-averse even for small-stake gambles and large initial wealth. This would predict a degree of risk aversion for high-stake gambles that is far away from standard behavior.

Other experimental evidence for systematic deviation from rational behavior has been accumulated over the last decades. One could joke that there is quite an industry for producing more and more such examples.

Does this mean, as is often heard, that the “homo economicus” is dead and that all models of humans as rational decision makers are obsolete? And does this mean that the excoriating judgment that we quoted at the beginning of this chapter holds in a certain way and that “science is at a loss” when it comes to people’s decisions?

Probably none of these fears is appropriate: the “homo economicus” as a rationally behaving subject is still a central concept, and on the other hand there are modifications of the rational theories that describe the irrational deviations from

the rational norm in a systematic way which leads to surprisingly good descriptions of human decisions. In the following we will introduce some of the most important concepts that such behavioral decision theories try to encompass.

The first example has already shown us one very important effect, the “framing effect”. People decide by comparing the alternatives to a certain “frame”, a point of reference. The choice of the frame can be influenced by phrasing a problem in a certain way. In Example 2.33 the problem was phrased in a way that made people frame it as a decision between *saving* 200 people for sure or saving 600 people with a probability of 1/3. In other words, the decision was framed in *positive* terms, in *gains*. It turns out that people behave risk-averse in such situations. This does not come as a surprise, since we have encountered this effect already several times, e.g., when we measured the utility function of a test person (see Sect. 2.2.4). In Example 2.34 the frame is inverted: now it is a decision about *letting people die*, in other words it is a decision about *losses*. Here, people tend to behave risk-seeking. They would rather take a 1/3 chance of letting all 600 persons die than choosing to let 200 people die.

But let us think about this for a moment. Doesn’t this contradict the observation that people buy insurances and that people buy lottery tickets? An insurance is surely about losses (and their prevention), whereas a lottery is definitely about gains, but still people behave risk-averse when it comes to insurances and risk-seeking when it comes to lotteries.

The puzzle can be solved by looking on the probabilities involved in these situations: In the two initial examples the probabilities were in the mid-range (1/3 and 2/3), whereas in the cases of insurances and lotteries the probabilities involved can be very small. In fact, we have already observed that lotteries which attract the largest number of participants typically have the smallest probabilities to win a prize, compare Example 2.21. If we assume that people tend to systematically *overweight* these small probabilities, then we can explain why they buy insurances against small probability risks and at the same time lottery tickets (with a small probability to win). Summarizing this idea we get a four-fold pattern of risk-attitudes¹⁴:

Can we explain Allais’ Paradox with this idea? Indeed, we can: When choosing between the lotteries A and B the small probability not to win anything when choosing A is perceived much larger than the difference in the probabilities of not winning anything when deciding between the lotteries C and D. This predicts the observed decision pattern (Table 2.5).

The fact that people overweight small probabilities should be distinguished from the fact that they often *overestimate* small probabilities: if you ask a layman for the

¹⁴It is historically interesting to notice, that a certain variant of the key ideas of Kahneman and Tversky have already been found 250 years earlier in the discussion on the St. Petersburg paradox: Nicolas Bernoulli had the idea to resolve the paradox by assuming that people *underweight* very small probabilities, whereas Gabriel Cramer, yet another Swiss mathematician, tried to resolve the paradox with an idea that resembles the value function of Prospect Theory.

Table 2.5 Risk attitudes depending on probability and frame

	Losses	Gains
Medium probabilities	Risk-seeking	Risk-averse
Low probabilities	Risk-averse	Risk-seeking

probability to die in an airplane accident or to get shot in the streets of New York, he will probably overestimate the probability, however, the effect we are interested in is a different one, namely that people even when they *know* the precise probability of an event still behave *as if* this probability were higher. This effect seems in fact to be quite universal, whereas the overestimating of small probabilities is not as universal as one might think. Indeed, small probabilities can also be underestimated. This is typically the case when a person neither experienced nor heard that a certain small probability event happened before. If you, for instance, let a person sample a lottery with an outcome with unknown, but low probability, then the person will likely not experience any such outcome and hence *underestimate* the low probability. Such a sampling will nowadays (in times of excessive media coverage) not be our only possibility to estimate the probabilities of events that we haven't experienced by ourselves before. But what about events that are too unimportant to be reported? Such events might nevertheless surprise us, since in these situations we have to rely on our own experience and tend to underestimate the probability of such events before we experience them. – Surely everybody can remember an “extremely unlikely” coincidence that happened to him, but it couldn't have been *that* unlikely if everybody experiences such “unlikely” coincidences, could it?

In the next section we formalize the ideas of framing and probability weighting and study the so-called “Prospect Theory” introduced by Kahneman and Tversky [KT79].

2.4.2 Original Prospect Theory

Framing effect and probability overweighting, these are the two central properties we want to include into a behavioral decision theory. We follow here the ideas of Kahneman and Tversky and use as starting point for this theory the Expected Utility Theory. Instead of the final wealth we consider the gain and loss induced by a given outcome (framing effect) and instead of the real probabilities we consider weighted probabilities that take into account the overweighting of small probabilities. This *Prospect Theory* (PT) leads us to the following definition of a “subjective utility” of a lottery A with n outcomes x_1, \dots, x_n (relative to a frame) and probabilities p_1, \dots, p_n :

$$PT(A) := \sum_{i=1}^n v(x_i)w(p_i), \quad (2.3)$$

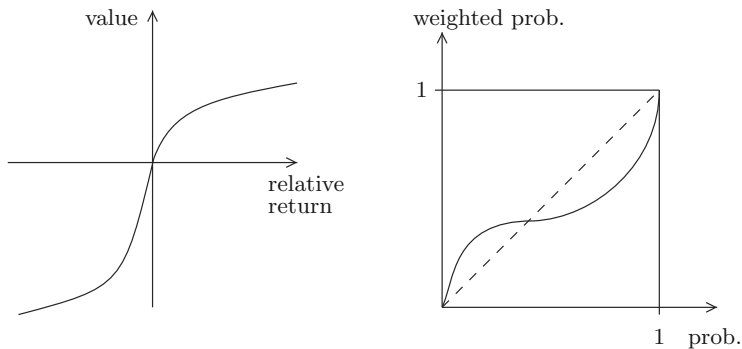


Fig. 2.13 A rough sketch of the typical features of value function (*left*) and weighting function (*right*) in Prospect Theory

where $v: \mathbb{R} \rightarrow \mathbb{R}$ is the *value function*, a certain kind of utility function, but defined on losses and gains rather than on final wealth, and $w: [0, 1] \rightarrow [0, 1]$ is the *probability weighting function* which transforms real probabilities into subjective probabilities. The key features of the value function are the following:

- v is continuous and monotone increasing.
- The function v is strictly concave for values larger than zero, i.e., in gains, but strictly convex for values less than zero, i.e., in losses.
- At zero, the function v is “steeper” in losses than in gains, i.e., its slope at $-x$ is bigger than its slope at x .

The weighting function satisfies the following properties:

- The function w is continuous and monotone increasing.
- $w(p) > p$ for small values of $p > 0$ (probability overweighting) and $w(p) < p$ for large values of $p < 1$ (probability underweighting), $w(0) = 0$, $w(1) = 1$ (no weighting for sure outcomes).

Typical shapes for v and w are sketched in Fig. 2.13.

If we have many events, all of them will probably be overweighted and the sum of the weighted probabilities will be large. There is an alternative formulation of Prospect Theory in [Kar78] that fixes the problem by a simple normalization:

$$PT(A) = \frac{\sum_{i=1}^n v(x_i)w(p_i)}{\sum_{i=1}^n w(p_i)}. \quad (2.4)$$

For didactical reasons it is easier to consider (2.3), hence we will mainly concentrate on this formulation. Equation (2.4) shares many of the common features with (2.3) and has some technical advantages that we will discuss later.

Can this new theory predict the four-fold pattern of risk-attitudes observed in the examples of the Sect. 2.4.1? Yes, it can. If we have two outcomes of similar probability, their weighted probability is approximately identical to their real probability, hence the concavity of the value function in gains, leads to risk-averse behavior, and the convexity of the value function in losses, leads to risk-seeking behavior. We know this already from EUT and do not need to compute anything new. (This explains the results of Examples 2.33 and 2.34.) Now if one of the probabilities is very small, then it is strongly overweighted ($w(p) > p$). In the case of losses this means that the overall utility is reduced even more. This effect can cancel the convexity of the value function and lead to a risk-averse behavior. On the other hand an overweighting of a gain might increase the value of the utility so much that it outperforms a sure option, even though the concavity of the value function would predict a risk-averse behavior.

Prospect Theory in this general form can only give a rough explanation of the experimental evidence, but is not useful for computations. To make precise predictions and to classify people's attitude towards risk, we need to make the functional forms of v and w precise. Nowadays the most commonly used functional forms are the ones introduced for *Cumulative Prospect Theory* (CPT), and we will discuss them in the next section. For the moment, we just note that Prospect Theory seems to be a good candidate for a descriptive model of decisions under risk. However, there are a couple of limitations to this theory that led to further developments.

We know that PT does not satisfy the Independence Axiom. This is a feature, not a bug, since otherwise we could not explain Allais' Paradox. There are some other axioms we are not so eager to give up in a descriptive theory. One of them is stochastic dominance: we have already briefly mentioned this concept which is essentially a "state-independent version" of state dominance:

Definition 2.36 (Stochastic dominance) A lottery A is *stochastically dominant* over a lottery B if, for every payoff x , the probability to obtain more than x is larger or equal for A than for B and there is at least some payoff x such that this probability is strictly larger.

This notion is quite natural: if we set our goal to get at least $x \in$ as payoff, we will choose A , since the probability to reach our goal with A is at least as large and sometimes strictly larger than with B . If this holds for all x , then A is in this sense "better" than B . If a decision criterion always prefers A over B when A is stochastic dominant over B , we say it *satisfies* or *respects stochastic dominance*.

Let us have a look at the following example: if we compare the two lotteries

$$A := \begin{array}{c|cccccc} \text{Outcome} & 0.95 & 0.96 & 0.97 & 0.98 & 0.99 & 1 \\ \hline \text{Probability} & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{array}, \quad B := \begin{array}{c|c} \text{Outcome} & 1 \\ \hline \text{Probability} & 1 \end{array},$$

then it is obvious that B is stochastic dominant over A (e.g., the probability to gain at least 0.97 is $1/2$ for A , but 1 for B) and should hence be preferred by a reasonable decision criterion. (Or would *you* prefer lottery A ?) In fact, it is easy to prove that EUT always satisfies stochastic dominance as long as the utility function is strictly increasing. Nevertheless, this does not need to be the case in Prospect Theory: the probability of $1/6$ is quite small, thus we expect $w(1/6) > 1/6$. On the other hand, the outcomes $0.95, \dots, 1$ are all quite close to 1, therefore

$$PT(A) = \sum_{i=1}^6 w(1/6)v(x_i) \approx \sum_{i=1}^6 w(1/6)v(1) > v(1). \quad (2.5)$$

This argument can easily be made rigorous to show that for every weighting function w that overweights at least *some* small probabilities, two lotteries can be constructed that show that PT violates stochastic dominance. In other words, if we want to have small probabilities being overweighted, there is no way we can at the same time rescue stochastic dominance. The alternative formulation (2.4) somehow reduces this problem such that stochastic dominance is not violated for lotteries with at most two outcomes, for lotteries with more outcomes, however, the problem persists. – This seems like bad news for the theory.

There is another problem involved in this example, namely a lack of continuity in this model. Roughly spoken, two very similar lotteries can have very different subjective utilities. We will discuss this problem in Sect. 2.4.5 more in detail.

Already Kahneman and Tversky knew about these problems and that their theory violates stochastic dominance and continuity. They suggested as “fix” a so-called “editing phase”: before a person evaluates the PT-functional (or rather behaves *as if* he evaluates this functional, since of course nobody assumes that people actually do these computations when deciding), this person would check a couple of things on the lotteries under consideration. In particular, the frame would be chosen, very similar outcomes would be collected to one, and stochastically dominating lotteries would automatically be preferred, regardless of any subjective utility.

The procedure is unfortunately not very well defined and leaves a lot of space for interpretations. (When are outcomes “close”? How does a person set the frame?) This causes problems when applying the theory and limits its usability.

Another limitation is that PT can only be applied for finitely many outcomes. In particular in finance, however, we are interested in situations with infinitely many outcomes. (An asset yields typically a return which can potentially be any amount, not just one out of a small list.)

We will discuss later why it is so difficult to extend PT to infinitely many outcomes and how one can improve PT regarding stochastic dominance and continuity (compare Sect. 2.4.6). Historically, however, these problems led first to the nowadays most important theory of behavioral decisions, the *Cumulative Prospect Theory*.

2.4.3 Cumulative Prospect Theory

We have seen that many problems in Prospect Theory were caused by the over-weighting of small probabilities. In a certain sense, our example for violation of stochastic dominance was based on the fact that a large number of small probability events added up to a “subjective” probability larger than one. The key idea of [TK92] was to replace the probabilities by *differences of cumulative probabilities*. In other words, we replace in the definition of Prospect Theory the probabilities p_i with the expression $F_i - F_{i-1}$, where $F_i := \sum_{j=1}^i p_j$ are the cumulative probabilities. (We set $F_0 := 0$.) Of course, the order of the events is now important, and we order them in the natural way, i.e., by the amount of their outcomes.

We write down the formula of Cumulative Prospect Theory precisely:

Definition 2.37 (Cumulative Prospect Theory¹⁵) For a lottery A with n outcomes x_1, \dots, x_n and the probabilities p_1, \dots, p_n where $x_1 < x_2 < \dots < x_n$ and $\sum_{i=1}^n p_i = 1$ we define

$$CPT(A) := \sum_{i=1}^n (w(F_i) - w(F_{i-1})) v(x_i), \quad (2.6)$$

where $F_0 := 0$ and $F_i := \sum_{j=1}^i p_j$ for $i = 1, \dots, n$.

There exist slightly different definitions of the CPT functional. In particular the original formulation in [TK92] differed in that it used the above formula only for losses, but a *de*-cumulative probability (i.e., $F_i := \sum_{j=i+1}^n p_j$) for gains. In finance, however, the above formula is more frequently used, since it is structurally simpler and essentially equivalent with the original formulation if one allows for changes in the weighting function.

How is this formula connected to Prospect Theory? Let us have a look on the case of a three-outcome lottery A (with outcomes x_1, x_2, x_3 with respective probabilities p_1, p_2, p_3) to see a little clearer, here the formulae reduce to

$$\begin{aligned} CPT(A) &= w(p_1)v(x_1) + (w(p_1 + p_2) - w(p_1))v(x_2) \\ &\quad + (1 - w(p_1 + p_2))v(x_3), \\ PT(A) &= w(p_1)v(x_1) + w(p_2)v(x_2) + w(p_3)v(x_3). \end{aligned}$$

We see that both formulae slightly differ, but not much. The difference between both models is essentially that in PT every probability is, regardless of their outcome, over- or underweighted, whereas in CPT, usually only probabilities that reflect

¹⁵The definition of CPT can be generalized if we use different weighting function w_- and w_+ for negative resp. positive outcomes. To keep things simple, we assume that $w_- = w_+ = w$.

extreme outcomes tend to be overweighted and probabilities that reflect outcomes in the middle are in general underweighted: if we compare the three terms in the formula for CPT, we see that the middle term indeed is likely to be the smallest, since the slope of w is typically small in the mid-range (compare Fig. 2.13). On average, events are neither over- nor underweighted in CPT¹⁶:

$$\sum_{i=1}^n (F_i - F_{i-1}) = F_n - F_0 = 1.$$

In many applied problems, probability distributions look similar to a normal distribution: extremely low and extremely high outcomes are rare, mid-range outcomes are frequent. This explains why often the difference between PT and CPT is small. Whereas PT overweights small probabilities which are associated with extreme outcomes, CPT overweights extreme outcomes which have small probabilities. Nevertheless, there can be situations where both theories deviate substantially, namely whenever small probability events in the mid-range of outcomes play a significant role.

There is another related theory, *Rank Dependent Utility* (RDU), which predates CPT and shares the cumulative probability weighting with CPT. However, it does not use the framing of PT and CPT, but uses a standard utility function in units of finite wealth, compare [Qui82].

In order to use CPT for applications, in particular in financial economics, we need to choose specific forms for v and w .

The prototypical example of a value function v has been given in [TK92] for $\alpha, \beta \in (0, 1)$ and $\lambda > 1$:

$$v(x) := \begin{cases} x^\alpha & , x \geq 0, \\ -\lambda(-x)^\beta & , x < 0, \end{cases} \quad (2.7)$$

compare Fig. 2.14. The parameter λ reflects the experimentally observed fact that people react to losses stronger than to gains: the resulting function v has a “kink” at zero, a marginal loss is considered a lot more important than a marginal gain. λ is usually assumed to be somewhere between 2 and 2.5.¹⁷

The probability weighting function w has been given by

$$w(p) := \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \quad (2.8)$$

¹⁶This is not the case in the original formulation of CPT when applying the weighting function on cumulative probabilities in losses and de-cumulative probabilities in gains.

¹⁷If $\alpha < \beta$, however, even a value of $\lambda < 1$ can lead to loss aversion. – In fact, when measuring λ on experimental data one often gets values substantially smaller than 2.

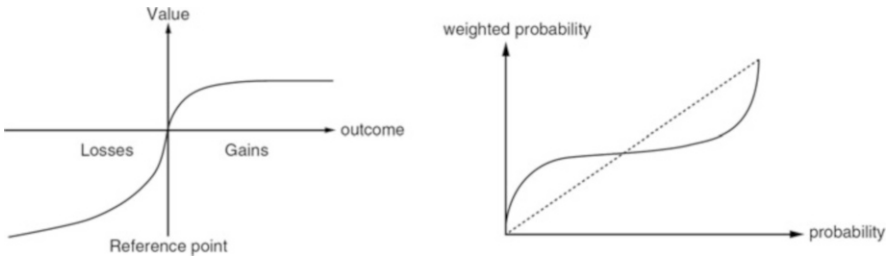


Fig. 2.14 Value and weighting function suggested by [TK92]

Table 2.6 Experimental values of α , β and γ , δ from various studies, compare (2.7) and (2.8) for the definition of α , β , γ , δ

Study	Estimate for α, β	Estimate for γ, δ
Tversky and Kahneman [TK92]		
Gains:	0.88	0.61
Losses:	0.88	0.69
Camerer and Ho [CH94]	0.37	0.56
Tversky and Fox [TF95]	0.88	0.69
Wu and Gonzalez [WG96]		
Gains:	0.52	0.71
Abdellaoui [Abd00]		
Gains:	0.89	0.60
Losses:	0.92	0.70
Bleichrodt and Pinto [BP00]	0.77	0.67/0.55
Kilka and Weber [KW01]	0.76–1.00	0.30–0.51

compare Fig. 2.14. It is possible to assign different weighting functions for gains and losses (denoted by w_+ and w_- , where in the loss region the constant γ is replaced by δ). There are also different suggestions how to choose v and w . We discuss in Sect. 2.4.4 which types of value and weighting functions are advantageous and which restrictions we have to take into account. For the moment we work for simplicity with the original suggestions by Tversky and Kahneman [TK92], although they are not the best choice. (For instance, w is not monotone increasing for $\gamma \leq 0.279$ [CH94, RW06, Ing07].) The parameters of their model have been measured experimentally in several studies, compare Table 2.6.

We see from this table that the results sometimes differ which might depend on the selection of the test sample or on the choice of the kind of experiment done to elicit these numbers. The overall impression, however, is that the values are typically in the range of $\alpha \approx \beta \approx 0.75 \pm 0.1$, $\gamma \approx \delta \approx 0.65 \pm 0.1$. Risk preferences also depend on economical and cultural factors, see [HW07a] for parameter estimates for some countries.

To fix ideas, we will in the following choose $\lambda := 2.25$, $\alpha := \beta := 0.8$ and $\gamma := \delta := 0.7$.

By the way, Prospect Theory (and also CPT) coincides with risk-neutral EUT when $\alpha = \beta = \gamma = \delta = \lambda = 1$. As can be seen from the experimental numbers, there is a strong deviation from this.

There are now two questions, we need to answer. Does the modified theory solve the problems that PT had (only finitely many outcomes, violation of stochastic dominance, lack of continuity) and does it still provide a good descriptive model of behavior under risk?

Let us first extend CPT to arbitrary lotteries. Since we all the time assume state-independent preferences, we can describe lotteries by probability measures, see Appendix A.4 for details.

Definition 2.38 Let p be an arbitrary probability measure, then the generalized form of CPT¹⁸ reads as

$$CPT(p) := \int_{-\infty}^{+\infty} v(x) \left(\frac{d}{dt} w(F(t))|_{t=x} \right) dx, \quad (2.9)$$

where

$$F(t) := \int_{-\infty}^t dp.$$

For the cognoscenti we remark that the formula (2.6) for lotteries with finitely many outcomes is just a special case of (2.9) when choosing p as a finite sum of Diracs.

Definition 2.38 paths the way to applications of CPT in financial economics and other areas where models require more than just a couple of potential outcomes. Although it looks at first glance much more involved than its finite counterpart (compare Definition 2.37), a closer look reveals the similarity: the sum in the definition of the cumulative probability is simply replaced by an integral, and the difference of weighted cumulative probabilities is replaced by a differential. Nothing special about this, it is just the usual process when proceeding from discrete to continuous situations.

We turn our attention now to stochastic dominance. Does CPT violate stochastic dominance? The answer is given by the following proposition:

Proposition 2.39 *CPT does not violate stochastic dominance, i.e., if A is stochastic dominant over B then $CPT(A) > CPT(B)$.*

¹⁸Here we consider again only the form defined in this book. In the original formulation we would need to write down two integrals for negative and positive outcomes and invert the direction of integration on the latter one. Compare the remark after Definition 2.37.

Proof We prove the case of finite outcomes. The general case is slightly tricky, in particular in the original formulation of CPT by Tversky and Kahneman, see, e.g., [Lév05].

Let x_i denote the potential outcomes of A and B . Let F_i denote the cumulative probabilities of A . Let G_i denote the cumulative probabilities of B . Then

$$\begin{aligned} CPT(A) &= \sum_{i=1}^n v(x_i)(w(F_i) - w(F_{i-1})) \\ &= \sum_{i=1}^n v(x_i)w(F_i) - \sum_{i=0}^{n-1} v(x_{i+1})w(F_i) \\ &= \sum_{i=1}^{n-1} (v(x_i) - v(x_{i+1}))w(F_i) + w(F_n)v(x_n). \end{aligned}$$

By Definition 2.36, we know that the probability to get a payoff of at most x_i with lottery A should be less or equal to the corresponding probability for lottery B . These probabilities are nothing else than F_i and G_i , and therefore we get $F_i \leq G_i$ for all $i = 1, \dots, n$ and that there is at least one i such that $F_i < G_i$. Moreover, using the monotonicity of v , $v(x_i) - v(x_{i+1}) < 0$. Finally $F_n = 1 = G_n$, so we get

$$\begin{aligned} CPT(A) &= \sum_{i=1}^{n-1} (v(x_i) - v(x_{i+1}))w(F_i) + w(F_n)v(x_n) \\ &> \sum_{i=1}^{n-1} (v(x_i) - v(x_{i+1}))w(G_i) + w(G_n)v(x_n) = CPT(B). \end{aligned}$$

This concludes the proof. \square

The final theoretical property that we hoped CPT to satisfy, since PT did not, is continuity. We expect that “similar” lotteries should have “similar” CPT values. The precise meaning of this will be explained in Sect. 2.4.5, for the moment we just convey that CPT is in fact continuous, compare Theorem 2.44. This excludes in particular any “event splitting effects”, in other words a lottery does not become more attractive if we partition an outcome into several very similar outcomes.

There is another attractive feature of CPT: it can be axiomatized. In other words, we can mimic the approach that von Neumann and Morgenstern used for Expected Utility Theory and define a set of axioms on preferences that describe Cumulative Prospect Theory. This has been observed first by Wakker and Tversky [Wak93]. Unfortunately, the axioms used are more complicated than in the case of Expected Utility Theory. The rough idea is first to replace the Independence Axiom with an equivalent set of (albeit less intuitive) axioms. This gives an alternative char-

acterization of EUT. Then one weakens this assumption by restricting the validity of these axioms only to a certain subclass of prospects. This characterizes CPT.¹⁹ By restricting the axioms to larger subclasses one also obtains two other decision models (Cumulative Utility and Sign-Dependent Expected Utility).

We have learned now that CPT is a conceptually adequate theory: it satisfies properties that we expect to hold for a behavioral theory for decisions under risk. Let us now take a look on the descriptive qualities of CPT. How well does CPT explain actual choices? Does it explain the phenomena we have encountered before as well as PT?

Let us first consider the Allais Paradox. If we choose v and w as the functions defined by Kahneman and Tversky (compare (2.7) and (2.8) for a definition) with the parameters $\lambda := 2.25$, $\alpha := \beta := 0.8$ and $\gamma := \delta := 0.7$, we can indeed explain the paradox by simply computing the CPT values of the four lotteries A, B, C and D. You may verify this as an exercise.

In general, we will also recover the four-fold pattern of risk-attitudes, but we have to change its definition slightly. Since we are not over- and underweighting solely depending on the size of the probabilities involved, things become a little bit more complicated. These complications, however, disappear as soon as we study the simple case of a lottery A with only two outcomes. The CPT functional in this case simply becomes

$$CPT(A) = w(p)v(x_1) + (1 - w(p))v(x_2).$$

Although this is not precisely the same formula as in PT,²⁰ it shares the same properties with it: small probabilities (either for x_1 or for x_2) are overweighted, large probabilities are underweighted. Since the value function has the same convex-concave shape in CPT as in PT, the four-fold pattern of risk-attitudes can be explained in exactly the same manner. – As long as we consider only two-outcome lotteries. This means in particular that we can explain the behavioral quirks that we encountered before: the life-death problem (Examples 2.33 and 2.34) and the fact that people both play lotteries and buy insurances.

We will see in the following chapters that CPT can also be used to explain several striking observations in finance, for instance the asset allocation puzzle. CPT has also been confirmed as a reasonable description of choices under risk by numerous quantitative studies.

After so much praise for this theory (which was a key reason for Daniel Kahneman to win the Nobel Prize in 2002 [TK92]), we also like to mention two limitations. To do so, we have to overcome a certain bias with which we have happily lived so far, namely that people are, if not fully, so at least partial, rational. We have tacitly assumed that people act according to the simple motto “more money is

¹⁹This characterization is mathematically quite involved. Brave readers might want to look into the original paper [Wak93].

²⁰Unless the weighting function w satisfies the symmetry property $w(1 - p) = 1 - w(p)$.

better” and apply the principle of stochastic dominance. Of course, one could always phrase a problem in a way that convinces people to make a wrong decision. (Some professions live from that.) But even if we provide clear, non-misleading conditions, this assumption, as natural as it seems, has been questioned severely in experiments. Let us have a look on the following example:

Example 2.40 There are two lotteries. In each case there are 100 marbles in total, one of which is drawn by chance. Every marble corresponds to a prize. The two lotteries have the following frequencies of marbles:

	Number of marbles	Prize in €		Number of marbles	Prize in €	
$A :=$	90	96	,	$B :=$	85	96
	5	14			5	90
	5	12			10	12

Which lottery do you prefer?

This example is taken from [BC97]. Which lottery did you choose? In several studies, a significant majority of persons preferred B over A . The percentage differed somehow with the educational background. (PhD students favored B only in around 50 % of the cases, whereas undergraduate students preferred in around 70 % of the cases.) What is wrong about this? You might have noticed that lottery A is stochastic dominant over B in the sense of Definition 2.36: the probability to win at least 96€ is larger for A , the probability to win at least 90€ is the same for both, the probability to win 14€ is again larger for A and in both cases you have the same probability (100 %) to win at least 12€, so in this sense, A really is better.

That A is stochastic dominant over B means in particular that not only EUT, but also CPT would predict a preference for A , since they both respect stochastic dominance. PT, however, can violate stochastic dominance, and in this particular case it can predict correctly that B is preferred over A . The reason for this difference is that PT overweights the intermediate outcome that occurs with only 5 % probability, but CPT does not. (Remember that CPT usually overweights only extreme events, not low probabilities in the mid-range.)

There are several other models for decision under risk that can predict such a behavior as well (e.g., RAM or TAX models, see [BN98, Bir05]), but since they are not used much in finance we refrain from describing them. Instead, we will give some information on a different way of extending PT that can describe this violation of stochastic dominance, but also allows for applications in finance (compare Sect. 2.4.6).

Important for us is to remember that we cannot expect people to follow *always* the stochastic dominance principle. Their decisions might deviate from this. This is not necessarily bad news, since deviations from rational behavior are, for instance, the key ingredients of active investment strategies! In *most* cases, however, assuming

that preferences are compatible with stochastic dominance is a safe thing to do, and it is enough to consider the irrational behavioral patterns like overweighting of small probabilities and framing effect that can be described well with PT or CPT.

2.4.4 Choice of Value and Weighting Function

When we use CPT (or PT) to model decisions under risk, we need to decide what value and weighting functions to choose. There are, in principle, two methods to obtain information on their shape: one is to measure them directly in experiments, the other one is to derive them from principal considerations. The former is the way that Tversky and Kahneman originally went, the latter one mimics the ideas that Bernoulli went with the St. Petersburg Paradox in the case of EUT.

Measuring value functions in experiments follow the same ideas outlined in Sect. 2.2.4. The measurement of the weighting function is more difficult. Some information on this can be found in [TK92] or [WG96]. The original choice of Kahneman and Tversky seems reasonable in both cases, although different forms for the weighting function have been suggested, the most popular being

$$w(F) := \exp(-(-\ln(F))^\gamma)$$

for $\gamma \in (0, 1)$, see [Pre98].

The measurement of these functions is of course limited to lotteries with relatively small outcomes. (Otherwise, laboratory experiments become too expensive.) This makes it also difficult to measure very small probabilities, since for small-stake lotteries, events with very small probability do not influence the decision much.

These are important restrictions if we want to apply behavioral decision theory to finance, since we will frequently deal with situations where large amounts of money are involved and where investment strategies may pose a risk connected to a very large loss occurring with a very small probability. We therefore are interested in finding at least some qualitative guidelines about the global behavior of value and weighting function based on theoretical considerations.

At this point it is helpful to go back to the St. Petersburg Paradox. We remember that the St. Petersburg Paradox in EUT was solved completely if we restricted ourselves to lotteries with *finite* expected value. Then the only structural assumption that we had to pose on the utility function was concavity above a fixed value.²¹ Does this result also hold for CPT? A closer look at this reveals some subtle difficulty: the far-out events of the St. Petersburg Lottery are overweighted by CPT which leads to a more risk-seeking behavior. (Remember the four-fold pattern of risk-attitudes!) Therefore one might wonder whether it is not possible to construct lotteries that have a finite expected return, but nevertheless an infinite CPT value.

²¹Compare Theorem 2.25 (ii).

This observation has been done in [Bla05] and [RW06]. The following result gives a precise characterization of the cases where this happens. We formulate it for general probability measures, but its main conclusions holds of course also for discrete lotteries with infinitely many outcomes.

Theorem 2.41 (St. Petersburg Paradox in CPT [RW06]) *Let CPT be a CPT subjective utility given by*

$$CPT(p) := \int_{-\infty}^{+\infty} v(x) \frac{d}{dx}(w(F(x))) dx,$$

where the value function v is continuous, monotone, convex for $x < 0$ and concave for $x > 0$. Assume that there exist constants $\alpha, \beta \geq 0$ such that

$$\lim_{x \rightarrow +\infty} \frac{u(x)}{x^\alpha} = v_1 \in (0, +\infty), \quad \lim_{x \rightarrow -\infty} \frac{|u(x)|}{|x|^\beta} = v_2 \in (0, +\infty), \quad (2.10)$$

and that the weighting function w is a continuous, strictly increasing function from $[0, 1]$ to $[0, 1]$ such that $w(0) = 0$ and $w(1) = 1$. Moreover assume that w is continuously differentiable on $(0, 1)$ and that there is a constant γ such that

$$\lim_{y \rightarrow 0} \frac{w'(y)}{y^{\gamma-1}} = w_0 \in (0, +\infty). \quad (2.11)$$

Let p be a probability distribution with $\mathbb{E}(p) < \infty$ and $\text{var}(p) < \infty$. Then $CPT(p)$ is finite if $\alpha < \gamma$ and $\beta < \gamma$. This condition is sharp.

In particular, the CPT value may be infinite for distributions with finite EV in the usual parameter range where $\alpha > \gamma$.

What does this tell us about CPT as a behavioral model? Did it fail, because it cannot describe this variant of the St. Petersburg Paradox? Fortunately, this is not the case: we can restrict the theory to a subclass of lotteries or we can change the shape of the value and/or weighting function. Roughly spoken, one can show that there are three ways to fix the problem [RW06]:

1. If we allow only for probability distributions with exponential decay at infinity (or even with bounded support), the problem does not occur. In many applications, this is the case, for instance if we study normal distributions or finite lotteries. However, in problems where we are interested in finding the optimal probability distribution (subject to some constraints), it might well happen that we obtain a “solution” with infinite subjective utility. This renders CPT useless for applications like portfolio optimization.
2. We could modify the weighting function w such that $w'(0)$ and $w'(1)$ are finite. This guarantees a finite subjective utility, independently of the choice of the value function (as long as it has a convex–concave structure).

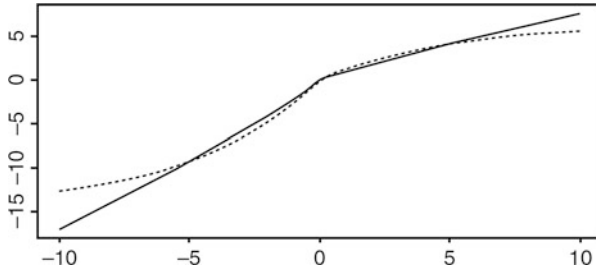


Fig. 2.15 Comparing classical (*solid line*) and exponential (*dashed line*) value function: they agree for small, but disagree for large outcomes

3. The value function can be modified for large gains and losses such that it is bounded. This again ensures a finite subjective utility. This is probably the best fix, since there are other theoretical reasons in favor of a bounded value function, compare Sect. 3.4.

There is of course a very strong reason in favor of keeping weighting and value function unchanged, namely that it has been introduced in a groundbreaking article and has subsequently used by many other people. Although this argument sounds strange at first, and arguments like this are often not fostering the scientific progress, there is in this case some grain of truth in it, namely that there is already a large amount of data on measuring CPT parameters, all based on the standard functional forms of value and weighting function. Changing the model means reanalyzing the data, estimating new parameters and generally making different studies less compatible.

How can we avoid such problems and still use functional forms that satisfy reasonable theoretical assumptions?

Fortunately, there are simple bounded value functions that are very close to the x^α -function used by Tversky and Kahneman, e.g. the exponential functions

$$v(x) := \begin{cases} \lambda^- e^{-\alpha x} - \lambda^- & , \text{ for } x < 0, \\ -\lambda^+ e^{-\alpha x} + \lambda^+ & , \text{ for } x \geq 0, \end{cases} \quad (2.12)$$

where the ratio λ^-/λ^+ corresponds to the loss aversion λ in PT and CPT, and α reflects the risk aversion (similar to PT and CPT). This function has been suggested in [DGHP05]. In Fig. 2.15 we compare the classical value function with the bounded variant. We see that the agreement for small values of x is very good. Since experiments are typically performed in this range, the descriptive behavior of both value functions should be very similar. For large values there is a strong disagreement which resolves the St. Petersburg Paradox and helps us applying CPT to problems in finance where we need a reasonable behavior of the CPT functional for lotteries involving the possibility of large gains and losses.

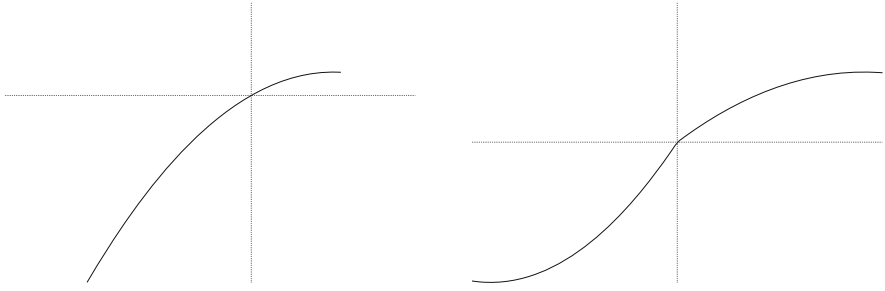


Fig. 2.16 A piecewise quadratic value function can describe MV- or PT-like preferences, depending on the parameters chosen

Another interesting example of an alternative value function has been introduced in [ZK09]: it makes an interesting connection between MV and PT by providing a common framework for both. Let us define the value function as

$$v(x) := \begin{cases} x - \alpha x^2 & , \text{ for } x < 0, \\ \lambda(x + \beta x^2) & , \text{ for } x \geq 0, \end{cases} \quad (2.13)$$

then for the case $\alpha = -\beta$ and $\lambda = 1$ we obtain a quadratic value function which implies that the corresponding decision model is the MV model – at least up to possible probability weighting and framing. By adjusting the parameters α , β and λ we can therefore generalize MV into the framework of PT which turns out particularly useful for applications in finance, compare Fig. 2.16. We will therefore use this functional form occasionally in later chapters.

There is, of course, the usual drawback in this specification that we inherit from PT and which is related to the mean-variance puzzle: the value function becomes decreasing for large values, thus we have to make sure that our outcomes do not become too large.

We have now developed the necessary tools to deal with decision problems in finance, from a rational and from a behavioral point of view. In the following section (which is intended for the advanced reader and is therefore marked with a *) we will discuss an interesting, but mathematically complicated concept in detail, namely continuity of decision theories. Afterwards, still marked with a * as warning to the nonspecialist, we introduce a different extension of PT that keeps more of the initial ideas of PT than CPT, but can nevertheless be extended to arbitrary lotteries. It might therefore be of some use for applications in finance, in particular in situations where the computation of CPT is computationally too difficult.

The nonspecialist might now turn his attention to Sect. 2.5, where we draw some connections between EUT, MV, PT and CPT. In particular, we will try to understand in which cases these theories agree, where they disagree, and in what situations we should apply them.

2.4.5 Continuity in Decision Theories*

We have already several times encountered the fundamental notion of *continuity*. This is a central property not only of decision theories, but virtually of all mathematical models, be it in economics, natural sciences or engineering. Its main insight is, that a model is only valuable if it allows for predictions that can be checked experimentally. In other words, for some given data, the model computes values for quantities that can be measured. Since the given data can in practical applications be never given with infinite precision, and it is also generally impossible to do computations with infinite accuracy, a fundamental property, which a reasonable model should satisfy, is that a slight change in the data only leads to a slight change in the predicted quantities. We call such behavior “continuous”.

Of course many systems are not continuous under every circumstances: think about the movement of a pendulum (a mass, attached with a bar to a fixed point) which can be predicted by the laws of gravity with high accuracy, unless we put the mass directly above the fixed point, from which a movement to either sides is equally likely and determined by indiscernibly small changes in the initial position (i.e., the data). However, in reasonable models such non-continuous situations should be a rare exception.

At this point, we need to add a word of warning: unfortunately, the word “continuous” has two quite different meanings in the English language: first, *continuous* means *non-discrete*. We have already used this notion when talking about measures (or lotteries), and we have seen how to extend the notion of EUT and CPT to such continuous distributions. Second, *continuous* means *not discontinuous*. This is what we mean when we speak about continuity in this section. Historically, both ideas are related, but nowadays they are distinct properties that should not be mixed up.

If we want to know whether a decision theory is continuous, we need to find a mathematically precise definition of continuity. In order to define continuity, we need to define what it means if $U(A_n) \rightarrow U(A)$, i.e., when the sequence of lotteries A_n converges to the lottery A . We know from calculus what it means if a sequence of numbers converges to another number, but what does it mean when lotteries converge? Intuitively, we would say that for $n \rightarrow \infty$, the following sequence of lotteries A_n converges to the lottery A with the certain outcome of 1:

$$A_n := \begin{array}{c} \text{Outcome} \quad 1 - \frac{1}{n} \quad 1 + \frac{1}{n} \\ \hline \text{Probability} \quad 1/3 \quad 2/3 \end{array}.$$

But how can we formulate this in mathematical terms? Fortunately, we can describe lotteries (in the state-independent setting which we consider here) by probability measures. There is a well-developed mathematical concept for the convergence of probability measures, but before giving the mathematical definition, we want to motivate it a little: we could say that a sequence of probability measures p_n converges to a probability measure p if every expected utility of p_n converges to

the expected utility of p . This would imply that no rational investor would see a difference between p and the limit of p_n . This idea leads to the mathematical concept of weak- \star -convergence:

Definition 2.42 (Weak- \star -convergence of probability measures) We say that a sequence $\{p_n\}$ of probability measures on \mathbb{R}^N converges weakly- \star to a probability measure p if for all bounded continuous functions f

$$\int_{\mathbb{R}^N} f(x) dp_n(x) \rightarrow \int_{\mathbb{R}^N} f(x) dp(x)$$

holds. We write this as $p_n \xrightarrow{\star} p$. The function f is sometimes called a *test function*.

To see the correspondence to the intuitive approach sketched above, we can consider $f(x)$ as a utility function.

In the above example, we can easily check that this definition is satisfied: first consider as f the indicator function²² on some interval $[x_1, x_2]$: in fact, if $x_2 < 1$, then the integral of A_n becomes zero when n is large enough, and the integral of A over this interval is also zero. The same holds if $x_1 > 1$. If $x_1 \leq 1 \leq x_2$ then the integral of A_n becomes eventually 1 and the integral of A is 1 as well. We then can approximate an arbitrary continuous function f by sums of indicator functions.

We can now formulate the definition of continuity²³:

Definition 2.43 (Continuity of a utility functional) We say that a utility functional U is *continuous*, if for all sequences of lotteries A_n with $A_n \xrightarrow{\star} A$ we have $U(A_n) \rightarrow U(A)$.

The concept of continuity, so natural it is in other situations, seems at first glance quite involved in the case of decision theory. However, having in mind that the mathematical formalism is just a way to clarify a quite intuitive concept (namely that “similar” lotteries should be evaluated in a “similar” way), is the main message we want you to remember.

Regarding the decision models we have encountered so far, we state that PT is discontinuous, whereas EUT, Mean-Variance Theory and CPT are continuous. We sketch a proof for the most complicated case, CPT. The other cases are left as an exercise for the mathematically inclined reader.

²²The indicator function is of course not continuous, but one can work around this problem by approximating the indicator function by continuous functions – a quite useful little trick that works here.

²³This definition is not related to the “Continuity Axiom” of von Neumann and Morgenstern (Axiom 2.18), even though the (unfortunate) name of the axiom suggests this.

Theorem 2.44 *If the weighting function w is continuously differentiable on $(0, 1)$ and the value function v is continuous, then CPT is weak- \star continuous.*

Proof We assume for simplicity that p is absolutely continuous. If $p_n \xrightarrow{\star} p$, then, by definition, $\int f dp_n \rightarrow \int f dp$ for all bounded continuous functions f . Using that p_n and p are probability measures and that p is absolutely continuous, one can prove that $F_n(x) = \int_{-\infty}^x dp_n \rightarrow \int_{-\infty}^x dp = F(x)$ for all $x \in \mathbb{R}$. Since w' is continuous, also $w'(F_n) \rightarrow w'(F)$. We compute

$$\int_{-\infty}^{+\infty} v(x) \frac{d}{dy} (w(F_n(y)))|_{y=x} dx = \int_{-\infty}^{+\infty} v(x) w'(F_n(x)) dp_n(x).$$

This is a product of a weak- \star converging term and a pointwise converging term. Using a standard result from functional analysis, this converges to the desired expression. \square

2.4.6 Other Extensions of Prospect Theory \star

Since we have seen that not all properties of CPT correspond well with experimental data (in particular its lack of violations of stochastic dominance), there are some descriptive reasons favoring PT. There is another, practical argument in favor of PT: computations in finance often involve large data sets and involved optimizations. In this case, PT is the computationally simpler model, since it does not need outcomes to be sorted by their amounts. For these reasons it is useful to look for an extension of PT to arbitrary (not necessarily discrete) lotteries. This is in fact possible if we use the variant of PT introduced by [Kar78], i.e.

$$PT(p) := \frac{\sum_{i=1}^n v(x_i) w(p_i)}{\sum_{i=1}^n w(p_i)}.$$

We assume as before that the weighting function w behaves for p close to zero like p^γ (with some $\gamma > 0$), compare (2.11).

The result of [RW08] is now summarized in the following theorem:

Theorem 2.45 *Let p be a probability distribution on \mathbb{R} with exponential decay at infinity and let p_n be a sequence of discrete probability measures with outcomes $x_{n,z}$ in equal distances of $1/n$ (each with probability $p_{n,z}$), i.e., $x_{n,z+1} = x_{n,z} + \frac{1}{n}$. Let $p_n \xrightarrow{\star} p$. Assume that the value function $v \in C^1(\mathbb{R})$ has at most polynomial growth and that the weighting function $w: [0, 1] \rightarrow [0, 1]$ satisfies the above condition. Then the normalized PT utility*

$$PT(p_n) := \frac{\sum_z w(p_{n,z}) v(x_{n,z})}{\sum_z w(p_{n,z})}$$

converges to

$$\lim_{n \rightarrow \infty} PT(p_n) = \frac{\int v(x)p(x)^\gamma dx}{\int p(x)^\gamma dx}.$$

This limit functional can therefore be considered as a version of PT for continuous distributions. A small problem is that we need to choose particular approximating sequences for p . Remark 2.50 shows how this can be fixed.

Theorem 2.45 can be generalized to lotteries that also contain singular parts. We summarize this in the following definition:

Definition 2.46 If p is a probability measure that can be written as a sum of finitely many weighted Dirac masses²⁴ $\pi_i \delta_{x_i}$ and an absolutely continuous measure p_a , i.e., $p = p_a + \sum_{i=1}^n \pi_i \delta_{x_i}$, then we can define

$$PT(p) := \frac{\sum_{i=1}^n v(x_i) \pi_i^\alpha + \int v(x) p_a(x)^\alpha dx}{\sum_{i=1}^n \pi_i^\alpha + \int p_a(x)^\alpha dx}.$$

Remark 2.47 The normalization is necessary, since otherwise the limit functional is either infinite (if $\gamma < 1$) or equivalent to a version of EUT (if $\gamma = 1$). Thus there would be no probability weighting in the limit.

Let us finally have a look at a related extension of PT [RW08]. Smooth Prospect Theory (SPT) encompasses parts of the editing phase of PT into the functional form, in that it collects “nearby” outcomes to one. This leads to a functional which is, unlike PT, continuous in the sense of the last section. We give here only its definition and some remarks on its properties:

Definition 2.48 Let p be a discrete outcome distribution. Then we define

$$SPT_\varepsilon(p) := \frac{\int w\left(\int_{x-\varepsilon}^{x+\varepsilon} dp\right) v(x) dx}{\int w\left(\int_{x-\varepsilon}^{x+\varepsilon} dp\right) dx}. \quad (2.14)$$

Remark 2.49 The parameter $\varepsilon > 0$ marks how small the distance between two outcomes can be until they are collected to one outcome. As long as $\varepsilon > 0$, SPT is continuous. It converges to PT when $\varepsilon \rightarrow 0$.

The definition of SPT allows us to generalize the convergence result of Theorem 2.45 to arbitrary approximating sequences:

²⁴For a definition of Dirac masses, see Appendix A.4.

Remark 2.50 If $p^k \xrightarrow{*} p$, then, for all sequences $k(\varepsilon) \rightarrow \infty$ that converge sufficiently slowly as $\varepsilon \rightarrow 0$, the SPT utility of p^k converges to $PT(p)$, i.e.:

$$\lim_{\varepsilon \rightarrow 0} SPT_{\varepsilon}(p^{k(\varepsilon)}) = PT(p) = \frac{\int v(x)p(x)^{\alpha} dx}{\int p(x)^{\alpha} dx}.$$

Proofs and further details on these results can be found in [RW08].

2.5 Connecting EUT, Mean-Variance Theory and PT

The main message of the last sections is that there are several different models for decisions under risk, the most important being EUT, Mean-Variance Theory and PT/CPT. The question we need to ask is: how important are the differences between these models? Maybe in “natural” cases all (or some) of these theories agree? In this section, we will check this idea. Moreover we will characterize the different approaches and their fields of applications. You should then be able to judge in a given situation which model is best to be applied.

First, we compare EUT and Mean-Variance Theory. Are they in general the same? Obviously not, since we have demonstrated in Theorem 2.30 that Mean-Variance Theory can violate state dominance, but we have seen in Sect. 2.2 that EUT does not, hence both theories cannot coincide. This shows that it is usually not possible to describe a *rational* person by Mean-Variance Theory.

This is certainly bad news if you still believed that Mean-Variance Theory is *the* way of modeling decisions under risk, but maybe we can rescue the theory by restricting the cases under consideration? This is in fact possible, and there are several important cases where Mean-Variance Theory can be interpreted as a special variant of EUT:

- If the von Neumann-Morgenstern utility function is quadratic.
- If the returns are all normally distributed.
- If the returns all follow some other special patterns, e.g., they are all lotteries with two outcomes of probability 1/2 each.
- In certain time-continuous trading models.

We will state in the following a couple of theorems that make these cases precise and show how they lead to an equivalence between both theories. First we define:

Definition 2.51 Let \succeq be an expected utility preference relation. We call EUT and Mean-Variance *compatible* if there exists a von Neumann-Morgenstern utility function $u(x)$ and a mean-variance utility function $v(\mu, \sigma)$ which both describe \succeq .

We have the following result:

Theorem 2.52 *Let \succeq be a preference relation on probability measures.*

- (i) *If u is a quadratic von Neumann-Morgenstern utility function describing \succeq , then there exists a mean-variance utility function $v(\mu, \sigma)$ which also describes \succeq .*
- (ii) *If $v(\mu, \sigma)$ describes \succeq and there is a von Neumann-Morgenstern utility function u describing \succeq , then u must be quadratic.*

Proof We prove (i): Let us write u as $u(x) = x - bx^2$. (We can always achieve this by an affine transformation.) The utility of a probability measure p is then

$$\begin{aligned} EUT(u) &= \mathbb{E}_p(u(x)) = \mathbb{E}_p(x - bx^2) = \mathbb{E}_p(x) - b\mathbb{E}_p(x^2) \\ &= \mathbb{E}(p) - b\mathbb{E}(p)^2 - b\text{var}(p) = \mu - b\mu^2 - b\sigma^2 =: v(\mu, \sigma). \end{aligned}$$

The proof of (ii) is more difficult, see [Fel69] for details and further references. \square

There is of course a problem with this result: a quadratic function is either affine (which would mean risk-neutrality and is not what we want) or its derivative is changing sign somewhere (which means that the marginal utility would be negative somewhere, violating the “more money is better” maxim) or that the function is strictly convex (but that would mean risk-seeking behavior for all wealth levels). None of these alternatives looks very appealing. The only case where this theorem can be usefully applied is when the returns are bounded. Then we do not have to care about a negative marginal utility above this level, since such returns just do not happen. The utility function looks then like $u(x) = x - bx^2$, $b > 0$, where $u'(x) > 0$ as long as we are below the bound. The minus sign ensures that $u'' < 0$, i.e., u is strictly concave. The drawback of this shape is that on the one hand it does not correspond well to experimental data and on the other hand there is no reason why this particular shape of a utility function should be considered as the only rational choice.

More important are cases where the compatibility is restricted to a certain subset of probability measures, e.g., when we consider only normal distributions:

Theorem 2.53 *Let \succeq be an expected utility preference relation on all normal distributions. Then there exists a mean-variance utility function $v(\mu, \sigma)$ which describes \succeq for all normal distributions.*

This means that, if we restrict ourselves to normal distributions, we can always represent an EUT preference by a mean-variance utility function.

Proof Let $N_{\mu,\sigma}$ be a normal distribution. Then using some straightforward computation and the substitution $z := (x - \mu)/\sigma$, we can define v :

$$\begin{aligned} EUT(u) &= \mathbb{E}_p(u(x)) = \int_{-\infty}^{\infty} u(x) N_{\mu,\sigma}(x) dx = \int_{-\infty}^{\infty} u(\mu + \sigma z) \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{-\infty}^{\infty} u(\mu + \sigma z) N_{0,1}(z) dz =: v(\mu, \sigma). \end{aligned} \quad \square$$

This idea can be generalized: the crucial property of normal distributions is only that all normal distributions can be described as functions of their mean and their variance. There are many classes of probability measures, where we can do the same. In this way, we can modify the above result to such “two-parameter families” of probability measures, e.g., to the class of log-normal distributions or to lotteries with two outcomes of probability 1/2 each.

After discussing the cases where Mean-Variance Theory and EUT are compatible, it is important to remind ourselves that these cases do *not* cover a lot of important applications. In particular, we want to apply our decision models to investment decisions. If we construct a portfolio based on a given set of available assets, the returns of the assets are usually assumed to follow a normal distribution. This allows for the application of Mean-Variance Theory as we have seen in Theorem 2.53. The assumption, however, is not necessarily true as we can invest into options and their returns are often not at all normally distributed. Given the manifold variants of options, it seems also quite hopeless to find a different two-parameter family to describe their return distributions.

We could also argue that the returns are bounded. Even if it is difficult to give a definite bound for the returns of an asset, we might still agree that there exists at least *some* bound. We could then apply Theorem 2.52, but this would mean that the utility function in the EUT model must be quadratic. Although theoretically acceptable, this seems not to fit well with experimental measurements of the utility function.

Finally, time-continuous trading is not the right framework in which to cast typical financial decisions of usual investors.

Therefore we see that there are many practical situations where Mean-Variance Theory does not work as a model for rational decisions. On the other hand, there are many situations where it is at least not too far from EUT (e.g., if the assets are not too far from being normally distributed etc.) and since Mean-Variance Theory is mathematically by far simpler than EUT, it is often for pragmatic reasons a good decision to use Mean-Variance Theory. However, results obtained in this way should always be watched with a critical eye, in particular if they seem to contradict our expectations.

How is it now with CPT (as prototypical representative of the PT family)? When does it reduce to a special case of EUT? How is its relation to Mean-Variance Theory?

Again, we see immediately, that CPT in general neither agrees with EUT nor with Mean-Variance Theory: it satisfies stochastic dominance, hence it cannot agree with Mean-Variance Theory, and it does not satisfy the Independence Axiom, thus it cannot agree with EUT.

How is it in the special case of normal distributions? In this case, the probability weighting does in fact not make a qualitative difference between CPT and Mean-Variance Theory, but the convex-concave structure of the value function can lead to risk-seeking behavior in losses, as we have seen. This implies that a person prefers a larger variance over a smaller variance, when the mean is fixed and contradicts classical Mean-Variance Theory.

We could also wonder how CPT relates to EUT if the probability weighting parameter becomes one, i.e., there is no over- and underweighting. In this case we arrive at some kind of EUT, but only with respect to a frame of gains and losses and not to final wealth. A person following this model, which is nothing else than the Rank-Dependent Utility (RDU) model, is therefore still not acting rationally in the sense of von Neumann and Morgenstern. We cannot see this from a single decision, but we can see this when we compare decisions of the same person for different wealth levels. There is only one case where CPT really coincides with a special case of EUT, namely when not only the weighting function parameter, but also the value function parameter and the loss aversion are one. In this case CPT coincides with a risk-neutral EUT maximizer, in other words a maximizer of the expected value.

On the other hand, we should not forget that CPT is only a modification of EUT. Therefore its predictions are often quite close to EUT. We might easily forget about this, since we have concentrated on the cases (like Allais' paradox) where both theories disagree. Nevertheless for many decisions under risk, neither framing effect nor probability weighting play a decisive role and therefore both models are in good agreement. We can illustrate this in a simple example:

Example 2.54 Consider lotteries with two outcomes. Let the low outcome be zero and the high outcome x million €. Denote the probability for the low outcome by p . Then we can compute the certainty equivalent (CE) for all lotteries with $x \geq 0$ and $p \in (0, 1)$ using EUT, Mean-Variance Theory, CPT. To fix ideas, we use for EUT the utility function $u(x) := x^{0.7}$ and an initial wealth level of 5 million €. For Mean-Variance Theory we fix the functional form $\mu - \sigma^2$ and for CPT we choose the usual function and parameters as in [TK92]. How do the predictions of the theories for the CE agree or disagree?

The result of this example is plotted in Fig. 2.17.

Summarizing we see that EUT and Mean-Variance Theory coincide in certain special situations; CPT usually disagrees with both models, but does often not deviate too much from EUT. We summarize the similarities and differences of EUT, Mean-Variance Theory and CPT in a diagram, see Fig. 2.18

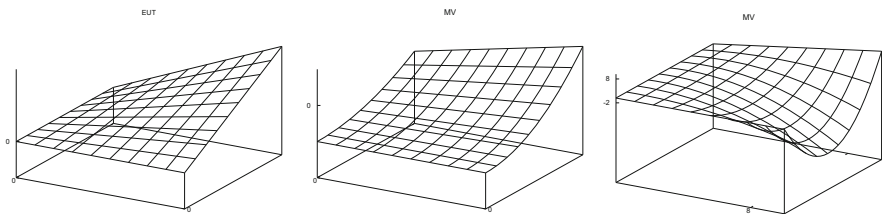


Fig. 2.17 Certainty equivalents for a set of two outcome lotteries for different decision models: EUT (*left*), CPT (*center*), Mean-Variance Theory (*right*). Small values for the high outcome x of the lottery are *left*, large values *right*. A small probability p to get the low outcome (zero) is on the back, a large probability on the front. The height of the function corresponds to its Certainty Equivalent

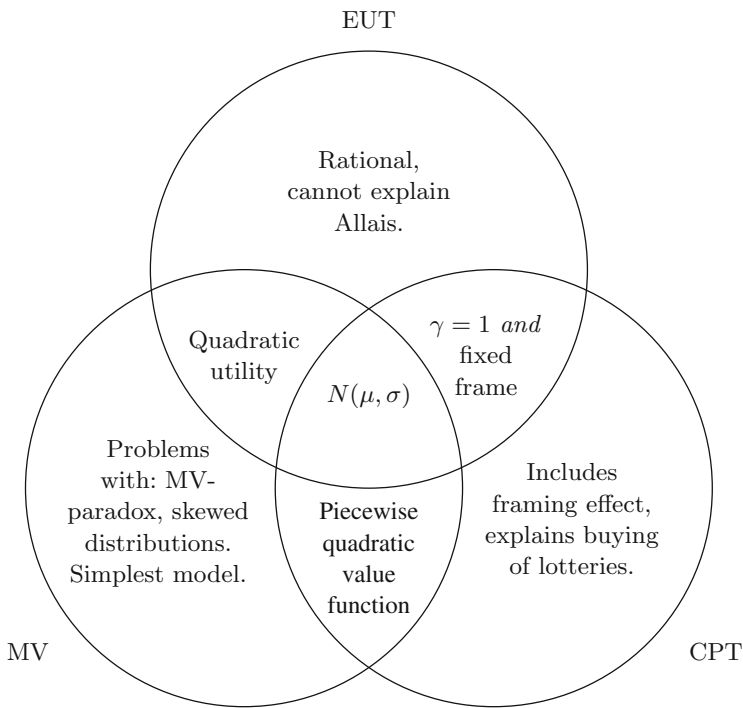


Fig. 2.18 Differences and agreements of EUT, PT and Mean-Variance

What does this tell us for practical applications? Let us sketch the main areas of problems where the three models excel:

- EUT is the “rational benchmark”. We will use it as a reference of rational behavior and as a prescriptive theory when we want to find an objectively optimal decision.

- Mean-Variance Theory is the “pragmatic solution”. We will use it whenever the other models are too complicated to be applied. Since the theory is widely used in finance, it can also serve as a benchmark and point of reference for more sophisticated approaches.
- CPT (and the whole PT family) model “real life behavior”. We will use it to describe behavior patterns of investors. This can explain known market anomalies and can help us to find new ones. Ultimately this helps, e.g., to develop new financial products.

We will observe that often more than one theory needs to be applied in one problem. For instance, if we want to exploit market biases, we need to model the market with a behavioral (non-rational) model like CPT and then to construct a financial product based on the rational EUT. Or we might consider the market as dominated by Mean-Variance investors and model it accordingly, and then construct a financial product along some ideas from CPT that is tailor-made to the subjective (and not necessarily rational) preferences of our clients.

In the next chapters we will develop the foundations of financial markets and will use all of the three decision models to describe their various aspects.

2.6 Ambiguity and Uncertainty*

We have defined at the beginning of this chapter that *risk* corresponds to a multitude of possible outcomes whose probabilities are known. Often we deal with situations where the probabilities are not known, sometimes they cannot even be estimated in a reasonable way. (What is the probability that a surprising new scientific invention will render the product of a company we have invested in useless?) In other occasions, there are ways to quantify the probabilities, but a person might not be aware of these probabilities. (Somebody who has no idea of the stock market will have no idea how (un)likely it is to lose half of his wealth when investing into a market portfolio, although a professional investor will be able to quantify this probability.) We call this *ambiguity* or *uncertainty*.²⁵

The difference between risk and uncertainty has first been pointed out by F. Knight in 1921, see [Kni21]. For the actual behavior of people, this difference is very important, as the famous Ellsberg Paradox [Ell61] shows:

Example 2.55 There is an urn with 300 balls. 100 of them are red, 200 are blue or green. You can pick red or blue and then take one ball (blindly, of course). If it is of the color you picked, you win 100€, else you don’t win anything. Which color do you choose?

²⁵Sometimes there are attempts in the literature to use both words for slightly different concepts, but so far there seems to be no commonly accepted definition, hence we take them as synonyms and will usually use the word “uncertainty”.

Which color did you choose? Most people choose red. Let us go to the second experiment:

Example 2.56 Same situation, you pick again a color (either red or blue) and then take a ball. This time, if the ball is *not* of the color you picked, you win 100€, else you don't win anything. Which color do you choose?

Here the situation is different: if you pick red, you win if either blue or green is chosen, and although you do not know the number of the green or the number of the blue balls, you know that there are in total 200. Most people indeed pick red.

However, this seems a little strange: let us say, in the first experiment you must have estimated that there are fewer blue balls than red balls, and hence picked red. Then in the second experiment you should have chosen blue, since the estimated combined number of red and green balls would be larger than the combined number of blue and green balls.

What happens in this experiment is that people go both times for the “sure” option, the option where they know their probabilities to win. In a certain way, this is nothing else than risk-aversity, but of a “second order”, since the “prizes” are now probabilities! One possible explanation of this experiment is therefore that people tend to apply their way of dealing with risky options, which works (more or less) well for decisions on lotteries,²⁶ also to situations where they have to decide between different probabilities. This is very natural, since these winning-probabilities can be seen as “prizes”, and it is natural to apply the usual decision methods that one uses for other “prizes” (being it money, honor, love or chocolate). Unfortunately, probabilities are different, and so we run into the trap of the Ellsberg Paradox.

It is interesting to notice that the “uncertainty-aversity” that we observed in the Ellsberg Paradox occasionally reverts to an uncertainty-seeking behavior, in the same way, the four-fold pattern of risk-attitudes can lead to risk-averse behavior in some instances and to risk-seeking behavior in others.

This is, however, only one possible explanation, and the Ellsberg Paradox and its variants are still an active research area, which means that there are many open questions and not many definite answers yet.

The Ellsberg Paradox has of course interesting implications to financial economics. It yields, for instance, immediately a possible answer to the question why so many people are reluctant to invest into stocks or even bonds, but leave their money on a bank account: besides the problem of procrastination (“I will invest my money tomorrow, but today I am too busy.”) which we will discuss in the next section, these people are often not very knowledgeable about the chances and risks of financial investments. It is therefore natural that when choosing between a known and an unknown risk, i.e., between a risk and an uncertain situation, they choose the safe option. This also explains why many people invest into very few stocks (that

²⁶We have seen that CPT models such decisions quite well, and that the rational decisions modeled by EUT are not too far away from CPT.

they are familiar with) or even only into the stock of their own company (even if their company is not performing well).

2.7 Time Discounting

Often, financial decisions are also decisions about time. Up to now we have not considered effects on decisions induced by time. In this little section we will introduce the most important notion regarding time dependent decisions, the idea of *discounting*.

A classical example for financial decisions strongly involving the time component is retirement, where the consumption is reduced *today* in order to save for *later*.

If you are faced with a decision to either obtain 100€ now or 100€ in 1 year, you will surely choose the first alternative. Why this? According to the classical EUT both should be the same, at least at first glance. On a second look, one notices that investing the 100€ that you get today will yield an interest, thus providing you with more than 100€ after 1 year. There are other very rational reasons not to wait, e.g., you may simply die in the meanwhile not being able to enjoy the money after 1 year. In real life, you might also not be sure whether the offer will really still hold in 1 year, so you might prefer the “sure thing”.

In all these cases, the second alternative is reduced in its value. In the simplest case, this reduction is “exponential” in nature, i.e., the reduction is proportional to the remaining utility at every time: if we assume that the proportion by which the utility u decreases is constant in time, we obtain the differential equation $u'(t) = -\delta u(t)$, where $\delta > 0$ is called *discounting factor*. This reduces the original utility $u(0)$ after a time $t > 0$ to

$$u(t) = u(0)e^{-\delta t}, \quad (2.15)$$

as we can see by solving the differential equation. If we consider only discrete time steps $i = 1, 2, \dots$, we can write the utility as $u(0)\delta^i$ (where the δ does not necessarily have the same value as before). To see this, set $t = 1, 2, \dots$ in (2.15).

Classical time discounting is perfectly rational and leads to a time-consistent preference: if a person prefers A now over B after a time t , this person will also prefer A after a time s over B after a time $s + t$ and vice versa:

$$\begin{aligned} u_B(s + t) - u_A(s) &= u_B(0)e^{-\delta(s+t)} - u_A(0)e^{-\delta s} \\ &= e^{-\delta s} (u_B(0)e^{-\delta t} - u_A(0)) \\ &= e^{-\delta s} (u_B(t) - u_A(0)), \end{aligned}$$

where we use that $e^{-\delta t}$ is a positive constant that does not influence the sign of the last expression.

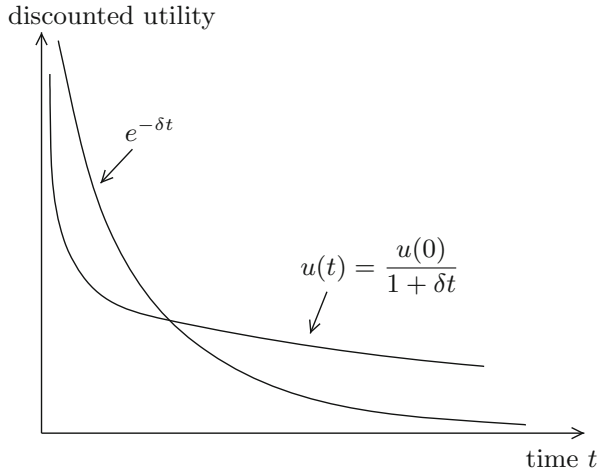


Fig. 2.19 Rational versus hyperbolic time discounting

Experience, however, shows that people do not behave according to the classical discounting theory: in a study test persons were asked to decide between 100hfl (former Dutch currency) now and 110hfl in 4 weeks [KR95]. Eighty-two percent decided that they preferred the money now. Another group, however, preferred 110hfl in 30 weeks over 100hfl in 26 weeks with a majority of 63 %. This is obviously not time-consistent and hence cannot be explained by the classical discounting theory. This phenomenon has been frequently confirmed in experiments. The extent of the effect varies with level of education, but also depends on the economic situation and cultural factors. For a large international survey on this topic see [WRH16].

The standard concept in economics and particularly in finance to model this behavior is the so-called “hyperbolic discounting”. The utility at a time t is thereby modeled by a hyperbola, rather than an exponential function, following the equation

$$u(t) = \frac{u(0)}{1 + \delta t}$$

where δ is the *hyperbolic discounting factor*, compare Fig. 2.19.

A similar definition is also often called hyperbolic discounting (or more accurately “quasi-hyperbolic” discounting), namely

$$u(t) = \begin{cases} u(0) & , \text{ for } t = 0, \\ \frac{1}{1+\beta} u(0) e^{-\delta t} & , \text{ for } t > 0, \end{cases} \quad \text{where } \beta > 0.$$

Hyperbolic discounting explains the behavioral pattern observed in the experiment by Roelofsma and Keren [KR95] and similar ones. Nevertheless, there is also some serious criticism against this concept, notably by Rubinstein [Rub03]

who points out that there are other inconsistencies in time-dependent decisions that cannot be explained by hyperbolic discounting, and that therefore the case for this model is not very strong. There is also recent work by Gerber [GR07] that demonstrates how uncertainties in the future development of a person's wealth can lead to effects that look like time-inconsistencies, but actually are not: in the classical experiment by Roelofsma and Keren, the results could e.g., be explained by classical time-discounting if people are nearly as unsure about their wealth level in the next week as in 30 weeks: the uncertainty of the wealth level reduces the expected utility of a risk-averse person at a given time. Although hyperbolic discounting is therefore not completely accepted, it is nevertheless a useful descriptive model for studying time-discounting.

A popular application of hyperbolic time-discounting is the explanation of undersaving for retirement. Here we give an example where hyperbolic discounting is combined with the framing effect:

Example 2.57 (Retirement) Assume a person has at time $t = 0$ a certain amount of money $w := 1$ which he could save for his retirement at time $t = 10$ yielding a fixed interest rate of $r := 0.05$. Alternatively, he can consume the interest rate of this amount immediately. The extra utility gained by consuming the interest rate wr is assumed to be wr and the utility gained by a total saving of x at the retirement age is $2x$, the factor 2 taking care of the presumably larger marginal utility at the retirement age, where the income, and hence the wealth level, shrinks. The hyperbolic discounting constant is $\delta = 0.25$. Does the person save or not?

We assume for simplicity that the person would either *always* or *never* save. A first approach would compare the discounted utility of the alternative “never saving” with the alternative “always saving”. A short computation gives

$$u(\text{always saving}) = \frac{u(w(1+r)^t)}{1 + \delta t} = \frac{2 \times 1.05^{10}}{3.5} \approx 0.9308,$$

$$u(\text{never saving}) = \frac{u(w)}{1 + \delta t} + \sum_{s=0}^t \frac{u(rw)}{1 + \delta s} \approx 0.8550.$$

This would imply that the person is indeed saving for his retirement. However, the decision whether or not to save might be framed differently: the person might decide on whether to start saving *now* or *tomorrow*. If he applies this frame²⁷ then his

²⁷This framing seems at least to be used frequently enough to produce proverbs like “A stitch in time saves nine” and “Never put off till tomorrow what you can do today”.

computation looks like this:

$$u(\text{start saving today}) = u(\text{always saving}) \approx 0.9308,$$

$$u(\text{start saving next year}) = \frac{u(w(1+r)^{t-1})}{1+\delta t} + u(wr) \approx 0.9365.$$

“Starting to save next year” is therefore the preferred choice – until next year, where the new alternative “starting to save yet another year later” suddenly becomes very appealing.

This theoretical explanation can also be verified empirically, e.g. by comparing data on time discounting from various countries with household saving rates [WRH16]: households in countries where people show stronger time discounting tend to save less.

The typical interaction of framing effect and hyperbolic discounting that we observe in retirement saving decisions can also be observed in other situations. Many students who start preparing for an examination in the last minute will know this all too well: one more day of procrastination seems much more preferable than the benefit from a day of hard work for the examination results, but of course everybody would still agree that it is preferable to start the preparation tomorrow (or at least *some* day) rather than to fail the exam. . .

2.8 Summary

Decisions under risk are decision between alternatives with certain outcomes which occur with given probabilities.

We have seen three models of decisions under risk: Expected Utility Theory (EUT) follows directly from the “rational” assumptions of completeness, transitivity (no “Lucky Hans”), continuity and independence of irrelevant alternatives (for a decision between A and B, only the differences between A and B matter). It is therefore the “rational benchmark” for decisions. The choice of the utility function allows to model risk-averse as well as risk-seeking behavior and can be used to explain rational financial decisions, e.g., on insurances or investments. The main purpose of EUT, however, is a *prescriptive* one: EUT helps to find the optimal choice from a rational point of view.

Sometimes EUT is too difficult to use. In particular when considering financial markets, it is often much easier to consider only two parameters: the expected return of an asset and its variance. This leads to the Mean-Variance Theory. We have seen that this theory has certain drawbacks, in particular it can violate state dominance. (This is called the “Mean-Variance paradox”.) In certain cases, in particular when the returns are normally distributed, Mean-Variance Theory turns out to be a special case of EUT, and hence we can more confidently use it.

EUT is about how people *should* decide. But how *do* people decide? The pessimistic statement of Chomsky on the unpredictable nature of human decisions,

which we had put at the beginning of this chapter, has been disproved to some extent in recent years: in particular Prospect Theory (PT) and Cumulative Prospect Theory (CPT) *describe* choices under risk quite well. Certain irrational effects like the violation of the “independence of irrelevant alternatives” make such approaches necessary to model actual behavior. Key features are the overweighting of small probabilities (respectively extreme events) and decision-making with respect to a reference point (“framing”). It is possible to explain the “four-fold pattern of risk-attitudes” and famous examples like Allais’ Paradox with these models.

Finally, we had a look on the time-dimensions of decisions. Whereas a discounting of the utility of future events can be explained with rational reasons, the specific kind of time-discounting that is observed is clearly irrational, since it is not time-consistent. Such time-inconsistent behavior can be used to explain, e.g., undersaving for retirement.

After finishing this chapter, we have now a very solid foundation on which we can build our financial market theories in the next chapters.

Tests

The following tests should enable the reader to check whether he understood the key ideas of decision theory. Some of the multiple choice questions are tricky, but most should be answered correctly.

Tests

1. How do you define that a lottery A with finitely many outcomes *state dominates* a lottery B with finitely many outcomes?
 - ☐ If A gives a higher outcome than B in every state.
 - ☐ If A gives a higher or equal outcome than B in every state, and there is at least one outcome where A gives a higher outcome than B .
 - ☐ If the expected return of A is larger than the expected return of B .
 - ☐ If, for every x , the probability to get a return of more than x is larger for A than for B .
2. What is the expected utility (EUT) of a lottery A with outcomes x_1 and x_2 and probabilities p_1 and p_2 ?
 - ☐ $EUT(A) = x_1p_1 + x_2p_2$.
 - ☐ $EUT(A) = u(x_1p_1 + x_2p_2)$.
 - ☐ $EUT(A) = u(x_1)p_1 + u(x_2)p_2$.
 - ☐ $EUT(A) = u(p_1)x_1 + u(p_2)x_2$.
3. Let us assume that u is an EUT utility function describing a person’s preference relation \prec , then:
 - ☐ $A \prec B$ if and only if $\mathbb{E}(u(A)) < \mathbb{E}(u(B))$.
 - ☐ $v(x) := u(2x + 42)$ is a utility function that describes the preference relation \prec .

- ☐ $v(x) := (u(x))^3$ is a utility function that describes \prec .
- ☐ If u is concave, then the person should not take part in any lottery that costs more than its expected value.
- ☐ If u is convex, then the person should take part in any lottery.
- ☐ If u is strictly convex on some interval then \prec cannot be rational.

4. In which cases is a function $u: [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ concave?

- ☐ If $\lambda u(x_1) + (1 - \lambda)u(x_2) \leq u(\lambda x_1 + (1 - \lambda)x_2)$ for every $x_1, x_2 \in [a, b]$ and $\lambda \in [0, 1]$.
- ☐ If $\lambda u(x_1) + (1 - \lambda)u(x_2) \geq u(\lambda x_1 + (1 - \lambda)x_2)$ for every $x_1, x_2 \in [a, b]$ and $\lambda \in [0, 1]$.
- ☐ If $\lambda u(x_1) + (1 - \lambda)u(x_2) = u(\lambda x_1 + (1 - \lambda)x_2)$ for every $x_1, x_2 \in [a, b]$ and $\lambda \in [0, 1]$.
- ☐ If $u'' \leq 0$.
- ☐ If $u'' \geq 0$.

5. The absolute risk aversion is defined by

- ☐ $r(x) := -u''(x)$.
- ☐ $r(x) := -u''(x)/u'(x)$.
- ☐ $r(x) := -x \frac{u''(x)}{u'(x)}$.

6. Which of the following utility functions is the most rational choice:

- ☐ $u(x) := x^\alpha$, where $\alpha \in (0, 1)$.
- ☐ $u(x) := x$.
- ☐ $u(x) := \ln x$.
- ☐ They are all equally rational.

7. What does Allais' Paradox tells us?

- ☐ It is irrational to follow Expected Utility Theory.
- ☐ Expected Utility Theory does not explain actual behavior of persons sufficiently well.
- ☐ People tend to violate the Independence Axiom.

8. Which are the key ideas of Prospect Theory (PT)?

- ☐ People frame their decisions in gains and losses rather than considering their potential final wealth.
- ☐ People tend to overweight small probabilities and underweight large probabilities. This can be modeled by a probability weighting function.
- ☐ People do not know probabilities exactly and hence overestimate small probabilities. This can be modeled by a probability weighting function.
- ☐ People compute the PT or CPT functional in order to make decisions.

9. How does PT explain why people gamble and buy insurances?

- ☐ People have a value function which is concave in gains (gamble) and convex in losses (insurance).
- ☐ People overweight small probabilities, like winning in a lottery or losing their home in a fire.

10. Why does PT violate stochastic dominance?

- ☐ Extreme events are overweighted, hence a small chance to lose a larger amount makes a lottery overly unattractive. This leads to a violation of stochastic dominance.
- ☐ Several small-probability events with similar outcome are overweighted relative to a single outcome with a slightly larger payoff, thus PT prefers the former to the latter, violating stochastic dominance.
- ☐ The convex shape of the value function in losses leads to risk-seeking behavior that makes people prefer risky lotteries over safe outcomes, violating stochastic dominance.

11. Which properties does Cumulative Prospect Theory (CPT) satisfy?

- ☐ Events with extremely low or high outcomes are overweighted.
- ☐ All small-probability events are overweighted.
- ☐ CPT does not violate stochastic dominance.
- ☐ CPT agrees with PT for lotteries with finitely many outcomes.
- ☐ CPT can be formulated for lotteries with finitely many outcomes as well as for arbitrary lotteries.

12. In which cases do Mean-Variance Theory and EUT coincide?

- ☐ When we consider only normal distributions of outcomes.
- ☐ When the utility function is concave.
- ☐ When the utility function is quadratic.
- ☐ When the utility function is linear.
- ☐ In lotteries with at most two outcomes.

13. Which axioms are satisfied by mean-variance theory?

- ☐ Completeness.
- ☐ Transitivity.
- ☐ Continuity.
- ☐ Independence.

14. *In-betweenness* says that the certainty equivalent of a lottery must be between its smallest and largest values.

Do the following four theories satisfy in-betweenness?

- ☐ Expected utility theory, i.e. $U = \sum p_i u(x_i)$,
- ☐ Classical prospect theory, i.e. $U = \sum w(p_i) v(x_i)$,
- ☐ Cumulative prospect theory, i.e. $U = \sum (w(F_i) - w(F_{i-1})) v(x_i)$,
- ☐ Normalized prospect theory by Karmarkar, i.e. $U = (\sum w(p_i) v(x_i)) / \sum w(p_i)$.

15. Which of the following statements on decision models are correct?

- ☐ From the von Neumann-Morgenstern axioms can we derive the existence of a utility function.
- ☐ A concave von Neumann-Morgenstern utility function corresponds to risk averse behavior.
- ☐ From the independence axiom we can derive that the utility function must be concave.
- ☐ Mean-Variance Theory describes rational decisions.
- ☐ EUT describes rational decisions.

- ☐ A typical utility function with constant relative risk aversion is $u(x) = x^\alpha / \alpha$.
- ☐ A typical utility function with constant relative risk aversion is $u(x) = -e^{-\alpha x}$.
- ☐ CPT is the most widely used descriptive model for decision behavior.
- ☐ Mean-Variance Theory can violate stochastic dominance.
- ☐ CPT can violate stochastic dominance.

16. Which of the following statements on time discounting are correct?

- ☐ In the classical model, the discounted utility at time $t > 0$ is given by $u(t) := \frac{u(0)}{1+\delta t}$ for some $\delta > 0$.
- ☐ In the classical model, the discounted utility at time $t > 0$ is given by $u(t) := u(0)e^{-\delta t}$ for some $\delta > 0$.
- ☐ Classical discounting is time-consistent, hyperbolic discounting is not.
- ☐ If somebody prefers 100€ now over 110€ tomorrow, this cannot be explained by classical discounting, but by hyperbolic discounting.
- ☐ If somebody prefers 100€ now over 110€ tomorrow, but 110€ in 101 days over 100€ in 100 days, then this cannot be explained by classical discounting, but by hyperbolic discounting.

Financial Economics

A Concise Introduction to Classical and Behavioral
Finance

Hens, T.; Rieger, M.O.

2016, XI, 363 p. 87 illus., 10 illus. in color., Hardcover

ISBN: 978-3-662-49686-2