Table of contents

Chapter 1: What decision analysis is about

1.0 Summary

1. Decision problems are often too complex to be tackled by common sense alone. Possible reasons for difficulties are uncertainty about the future or a variety of conflicting objectives. Other causes might be the existence of too many or too few alternatives, or an overwhelming number of influencing factors.
2. A structured and rational process can help improve the chances of receiving good decision outcomes. Rationality is not a clear-cut concept. It can, however, be put into more concrete terms by defining some procedural requirements (procedural rationality) and agreeing on some basic claims regarding consistency.
3. A complex decision problem can be simplified by decomposing it into its most basic components.
4. Objectively correct decisions do not exist. In fact, decisions are necessarily based on subjective expectations, which are often difficult to validate, as well as subjective objectives and preferences of the decision maker.
5. Articulating exact expectations and preferences is not an easy task for most people. Decision analytic tools and procedures have to take into account the fact that the information collected is often "fuzzy". Frequent consistency checks and specific techniques that allow the handling of incomplete information are two approaches that are used in decision analysis to cope with this problem.
6. *Prescriptive* decision theory (decision analysis) is of great relevance to many areas of application. It is not only important for economic decisions but also for political, medical, legal, or technical decision making. Furthermore, decision analytic thinking can be helpful for many personal and daily life decisions.

1.1 Why is decision making difficult?

This text book is an introduction to *prescriptive* decision theory, also known as *decision analysis*. Decision analysis aims to support people in making complex and hard decisions.

In contrast, the goal of *descriptive* decision theory is to describe how people actually behave when making decisions. Despite these different objectives, the insights provided by descriptive decision theory are very important for decision analysis. Psychologists have revealed a number of systematic errors to which we are all susceptible in our daily decision making. These errors not only often lead to bad intuitive decisions; they might also constrain the application of decision ana-

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_1, © Springer-Verlag Berlin Heidelberg 2010

lytic procedures. Because of this, we will frequently address descriptive issues in this book about prescriptive decision theory.

The starting point of decision analysis is the insight that humans have trouble when confronted with unfamiliar, non-routine decision situations. Why is decision making so difficult?

First of all, there is *uncertainty* about the future. It cannot be stated for sure what consequences will result from the choice of one or another alternative. For a patient in hospital, having to decide whether or not to undergo surgery that in the worst case scenario could lead to her death, uncertainty is surely the crucial problem. A similar dilemma is faced by a plaintiff who is unsure whether his compensation claim of €200,000 will be successful in court. The defendant is offering him €100,000 as an out-of-court settlement. Should he accept the proposed settlement or run the risk of engaging in the lawsuit with the possible outcome of not receiving anything at all? Many research and development projects and decisions about bringing new products to the market are strongly impacted by uncertainty.

As a second important aspect, the consequences of different actions can vary in more than one dimension. The choice between a station wagon and a sports car can turn out to be a difficult one, as the decision maker has multiple objectives – he cares about horsepower, passenger comfort, vehicle safety, styling, and many other attributes of the car, hence resulting in a conflict of interest. This problem occurs in many relevant decision situations. In recruiting decisions, various characteristics of the applicants normally have to be considered: work experience, education, reliability, leadership qualities, negotiating skills, etc. For the decision concerning the best location for a factory, numerous objectives such as reasonable property costs, accessibility, environmental constraints and the local job situation, among others, can be relevant. Conflicting objectives are often a problem for an individual decision maker, and they become even more problematic in situations where many decision makers are involved.

A third potential difficulty arises if too few or too many *alternatives* are available. In the first case, the dilemma is that either additional effort has to be invested into the search for additional alternatives – and success is not guaranteed – or that one of the given, marginally attractive options has to be chosen. In the second case, pre-selection strategies are needed to reduce the abundance of alternatives to a manageable set of options that can be considered in more detail.

The complexity of a decision situation increases if more factors have to be taken into account. The more sources of uncertainty have an impact on outcomes, and the more objectives need to be considered, the more complex the situation becomes. Whether or not to buy home insurance is a relatively simple decision problem. In contrast, political decisions about how to combat an approaching global recession are very complicated. Because of the many interests that have to be considered and the uncertainty about the consequences, many political decisions are extremely complex. Should we rely further on nuclear power or strive for alternative types of energy production and subsidize them? Should stem cell research be supported or prohibited? Should a damaged banking system be backed by public money, shifting the burden to tax payers? Would it be a sensible decision to legalize (in a controlled way) all types of drugs?

The more complex a decision situation becomes, the more important it is to support it with procedures and tools that have been developed for a systematic elicitation and processing of relevant information. Humans need formal rules and procedures to shape and clarify expectations and objectives that are often fuzzy, unclear and contradictory. They also need such procedures to process information consistently. Keeney (1982) defines decision analysis as "a formalization of common sense for decision problems which are too complex for informal use of common sense". In the words of Howard (1988):

> Today we have a discipline of decision analysis: a systematic procedure for transforming opaque decision problems by a sequence of transparent steps. Opaque means "hard to understand, solve, or explain; not simple, clear, or lucid." Transparent means "readily understood, clear, obvious." In other words, decision analysis offers the possibility to a decision maker of replacing confusion by clear insight into a desired course of action.

The very general approach of decision analysis ensures its applicability in various decision scenarios, ranging from the highly complex problem of picking a location for a major new airport to the relatively simple choice of a new Blu-ray player. In various areas of our society, there is a need for constant decision making with high relevance for many people: in politics, companies, agencies, schools and universities, hospitals, and courtrooms. When observing such important decision situations, one is occasionally puzzled about the lack of structure and systematic analysis in the decision making process. The same holds for many important personal decisions that are often founded on intuition rather than on structured analysis and a methodical dissection of the problem.

The special way of thinking proposed by decision analysis can be helpful in any decision situation. It should thus be a mandatory part of education and training for any profession that requires frequent decision making. It is not only relevant to managers but also to physicians, politicians, judges and engineers. Moreover, situations of helplessness and indecisiveness also occur frequently in people's personal lives. Here, structured analytical thinking can help make decisions perceived as rational and well-founded.

1.2 Basic principles of decision analysis

1.2.1 Striving for rationality

The goal of decision analysis is to help decision makers make the most rational decisions that is possible. Unfortunately, there is no clear-cut definition of the term "rationality", making it difficult to evaluate the quality of a particular decision. One should not speak of "rational" and "irrational" but rather of "more or less rational". It may seem that the quality of a decision can, at the very least, be evaluated retrospectively by considering the outcome. Ultimate success or failure is, however, not a sound measure. If you decided to buy a specific stock after a

thorough analysis and its price happened to decrease afterwards, this does not necessarily mean that the original decision was irrational. If an impoverished college student decides to gamble in a casino, puts all his remaining money on number 17 and happens to win, his decision does not become more rational in retrospect – he was simply lucky and won. There is an important distinction to be made between rational and successful decisions – of course, the aim of increased rationality is to make more successful decisions. However, if outcomes are uncertain, one can have good or bad luck. Decisions that occur frequently or are less uncertain can more effectively be evaluated in terms of their outcomes than can unique decisions that have to be made under extreme uncertainty. If a material requirement planner causes shortages of raw materials repeatedly, one can assume that he frequently makes bad decisions. If a surgeon, by contrast, decides that a seriously ill patient should undergo surgery and the patient dies in the process, this tragic outcome does not indicate that the original decision was irrational.

Retrospective success control is nevertheless important as it can help uncover problems in the decision making process. If, for example, the launch of a new product turns out to be a disaster, it is worth considering if we thought about the possibility of demand being this low? Did we produce a realistic assessment of the riskiness of the launch? If not, would it have been possible to have known better, had a more thorough analysis been performed? Such retrospective evaluation will only be insightful, however, if all components of the decision making process are fully and transparently documented. Producing optimal transparency is one of the major goals of decision analysis. Thorough documentation will also help avoid hindsight bias. Hindsight bias refers to people's tendency to assume retrospectively (i.e. when the outcome of the decision is known) that they would have expected such an outcome also from an ex ante perspective, or that someone else should have expected it. This type of judgment bias can lead to wrongly blaming someone else for one's own bad decision in the case of failure, or to praising oneself overly for excellent decision making in the case of success.

What does "rational" hence mean? It is not an objective and provable property. We can do no better than to specify some basic requirements that seem plausible to most decision makers. However, it is up to each and every one of us to accept these requirements or to renounce them. If you believe that you will do best by just following your gut feeling, or if you think that how things will evolve is predetermined by fate anyway, you do not need decision analysis. (However, the authors of this book would certainly refrain from voting for a politician who receives his advice from a fortune teller and her crystal ball).

Striving for rationality does not provide any guarantee of successful decisions. Presumably, it can help one become a more successful decision maker on average; this, however, cannot be proven.[1]

In the following discussion, we will distinguish between two criteria for rational decision making. The first is called *procedural rationality*, the second one *consistency*. The latter concept refers to the consistency of assumptions in the deci-

[1] See Rescher (1993), Chapter 3.

sion scenario as well as to compliance with generally accepted rules of rational thinking.

1.2.2 Procedural rationality

The *procedure* that leads to the final decision can be more or less rational. The following section specifies some of the basic guidelines that a decision procedure should adhere to:

1. The decision maker has to make sure that he is tackling the *right problem*. Each decision refers only to a subset of all the problems that a decision maker has to solve throughout his life. Individuals and also organizations tend to solve problems by a patchwork approach, changing the status quo as little as possible. However, it could be advisable to extend the scope of the decision problem to look for a global rather than a local optimum for the problem at hand. Conversely, it might be helpful to split a large and complex problem into one "sub-problem" that has to be solved now and some remaining sub-problems that can be delayed until some future point.
2. The decision maker should invest as much *effort* into the collection and processing of information as is appropriate and of relevance to the decision. Rationality calls for a preparation for the decision that is thorough and systematic, but not necessarily *maximized* – otherwise, decision problems would never be solved. Simplification is indispensable.
3. The decision maker should take into account relevant objective data when forming *expectations* about the future. He should be aware of the existence of judgment bias and pay special attention to avoiding it.
4. The decision maker should think carefully about his *objectives* and *preferences*. He should be aware of potential problems of self-deception and limited powers of imagination and take action to deal with these shortcomings.

Some examples:

- *The right problem.* A company has recognized that the quality of one of its products is insufficient and is considering different measures to improve quality control, e.g. extended testing or quality-dependent premiums. All these measures produce costs and the company tries to find the solution that promises to ensure the desired quality at the lowest cost. Could it be that the problem is too narrowly framed? What about a much broader, more strategic approach? Perhaps this would make solutions such as "accepting the low quality and decreasing the price" or "outsourcing the production of specific critical components" viable alternatives.
- *Appropriate information effort.* In general, the more there is at stake and the more complex the problem, the more valuable the effort of collecting and evaluating information becomes. Some people spend hours on their hunt for the cheapest asparagus in town but decide spontaneously – and without looking for alternatives – to buy a leather jacket just because it is marked down from €1,700 to €700. Humans evaluate prices relative to a benchmark – if they want to buy a radio for their old car, a €200 difference in price between two possible alternatives leads to pondering and reflection, but when buying a new car, they easily make the decision to pay an additional €500 for the more exclusive radio because, with all the other expenses, "it doesn't matter anymore anyway".
- *Forming expectations.* In decision analysis, uncertainty about the future is captured by *probabilities*. What facts should be taken into account when deriving probabilities? The owner of an apple plantation has to decide whether he should protect his apple trees from night frost. This decision requires him to evaluate the probability of night frost occurring within the next month. He could evaluate historical data on night temperatures during that time of the year; he could call the meteorological service or trust the local folk sayings. Most of us will agree that historical weather data and official forecasts are a more rational basis for decision making than are folk sayings.
- *Forming objectives and preferences.* If we face novel decision situations, we often have no clear and complete idea of our objectives and preferences. We tend to concentrate on a few or even just a single attribute of the alternatives. When reflecting on the next car we want to buy, the price might look most relevant right now but driving comfort seems most important after a test drive tomorrow and after reading a test report the following day, safety issues seem essential. Not until the decision maker has compiled a complete set of relevant aspects and has arrived at a robust judgment concerning their relative importance can we consider his decision to be truly rational.

It will never be possible to determine objectively and accurately to what extent these properties of rational decision making are fulfilled; nevertheless, it is important to align the decision making process in their direction.

1.2.3 Consistency of the decision foundation

Rational decision making cannot be based on assumptions that are mutually contradictory. It is not reasonable to consider astrology to be a superstition but at the same time consider the signs of the zodiac when choosing a partner. Rationality further requires conformity with specific standards which the decision maker accepts and aims to meet. Such "standards of rationality" refer on the one hand to the processing of probabilities and on the other to the process of forming preferences.

Below is a famous example of the intuitive violation of basic rules of probability calculus (Tversky and Kahneman 1982):

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of the following is more likely:
>
> (a) Linda is a bank teller.

(b) Linda is a bank teller and is active in the feminist movement.

Many people intuitively think that (b) is more likely than (a). However, this is at odds with basic rules of probability theory as (b) is a sub case of (a).

Rational preferences are usually expected to fulfill the following major requirements:

1. Prospective orientation. A choice between alternatives should only be dependent on the respective consequences.
2. Transitivity. If a decision maker prefers a to b and b to c then he should also prefer a to c.
3. Invariance. Preferences should not be influenced by the mere framing of the decision problem, given the descriptions are equivalent.
4. Independence from irrelevant alternatives. Whether a decision maker prefers a or b should not be influenced by the existence or non-existence of a third alternative c.

Prospective orientation

The principle of prospective orientation requires rational decisions to rely only on aspects that are influenced by the decision and thus lie in the future. In contrast, it is irrational to let a decision be influenced by past events that cannot be changed anymore. In this spirit, the following decision pattern must be considered irrational: an investor is reluctant to sell some of his stocks at the current price of €300 per share. He argues that he would like to "avoid the loss that he would incur given that he once bought the stocks at a price of €350 per share". When checking his files, he realizes that he was wrong and he actually bought the stocks at a price of €280 per share. Relieved, he now places the order to sell. Such a decision pattern violates the principle of prospective orientation. The decision to sell should only depend on expectations for the future and not on the original purchase price of the stock.

Another interesting example shows up in the reasoning of the former US President George W. Bush. He presented different justifications for a continuation of the war in Iraq, despite the unsatisfactory progress and the fact that one original rationale for war, the assumed existence of weapons of mass destruction, had been refuted. One of his new arguments was that many US soldiers had died in the course of the war. "We owe them something," he told veterans. "We will finish the task that they gave their lives for." (see e.g. the column in the New York Times from August 24th 2005). One possible interpretation of this statement which is useful for purposes of our discussion is that the previous victims call for a continuation of the war. This argument would violate the postulate of prospective orientation.

The past cannot be changed by the decision made now. Bygone and irreversible investments are referred to as *sunk costs* and it is a common decision error to take them into account in the decision making process. When observing such reasoning and decision making in a politician, one should be careful not to draw the immediate conclusion that he personally lacks rationality. For most politicians, the system of objectives is much more complex than communicated to the public. Additionally, the presented rationale for the decision making might just be used as a pretext – it might also be the case that President Bush is well aware of the irrationality of the reasoning, but that he knows at the same time how convincing such an argument sounds to most citizens. It is nevertheless important to understand that the logic that the existence of previous victims justifies the continuation of the war is very problematic from a rational point of view.

Transitivity

If you prefer fish to burgers and burgers to pudding, you should also prefer fish to pudding. However, we do not always adhere to this principle, as shown in the following example:

Professor P is looking for a new job. She knows that large salary differences between two options will make her focus on this particular attribute. If salary differences are moderate, other factors such as the reputation of the university will gain more relevance. She ends up with three offers that can be described partially in the following way:

	Salary	Reputation
x	\$65,000	Low
y	\$50,000	High
z	\$58,000	Mediocre

After some reflection, P realizes that she prefers x to y, y to z, and z to x (Fishburn 1991, p. 120).

Invariance

Consequences of decisions can often be presented in different ways. The efficiency of a vaccination against an epidemic can be measured by the percentage of people saved by the vaccination, but also as the percentage of people that died despite the vaccination. If two presentations of a decision problem are equivalent, i.e. one can be transformed into the other, the invariance postulate requires the decision not to be influenced by the form of presentation. A familiar example taken from Tversky and Kahneman (1981), the so-called "Asian disease problem", demonstrates that people often violate this principle in intuitive decision making:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. There are two alternative programs.

1. If Program A is adopted, 200 people will be saved. If Program B is adopted, there is a one-third probability that all 600 people will be saved and a two-thirds probability that no one will be saved. Which do you prefer, Program A or Program B?

In this decision situation, most people choose program A. The choice pattern is very different, however, if the same problem is described in a slightly different way and (1) is replaced by:

2. There are two alternative programs. If Program A is adopted, 400 people will die. If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 will die. Which do you prefer, Program A or Program B?

Given this description, a majority of decision makers choose program B, a clear violation of description invariance.

Independence from irrelevant alternatives
The meaning of this requirement can be illustrated by the following example. The menu in a restaurant offers two dishes: salmon and schnitzel. When ready to order the salmon, you are informed by the waiter that they are also offering pork chops with sauerkraut today. After some reflection on the three options, you decide to order the schnitzel. Such a decision pattern is irrational – the "irrelevant" alternative of pork chops should not influence the preference you have with respect to salmon and schnitzel.

These requirements might seem more or less self-evident. Nevertheless, they are violated so frequently – even in simple decision situations – that they cannot serve as descriptive hypotheses of human behavior. The fact that intuitive behavior so often violates the most basic rationality principles increases the relevance of decision analysis for all people eager to obey the rules of rationality. Nonetheless, it should also be mentioned that not all decision theorists agree with our claim that all the abovementioned principles are indeed necessary for rational decision making.

1.2.4 Decomposition

The complexity of a decision problem can be reduced by dividing it into its basic components. These basic components that have to be modeled are:

1. the alternatives (or actions) which the decision maker has to choose from,
2. the decision maker's objectives and preferences that are relevant in the decision context,
3. expectations with respect to uncertainty (i.e. events that cannot be changed by the decision maker),
4. the combined impact of actions and uncertainties on the outcome of the decision, i.e. consequences of the decision.

The following simple example shows how these components can be extracted from the description of a decision problem: A manager has to attend an important meeting at 9 a.m. on Monday morning in Cologne. He lives in Aachen and has to decide how to get to Cologne. His analysis of the problem produces the following components.

The most obvious *alternatives* are: travelling by train from Aachen to Cologne on Monday morning, driving to Cologne by car on Monday morning or driving to Cologne by car on Sunday evening and spending the night at a hotel.

The *objectives*: It is important to the manager not to arrive late at the meeting; however, this is not the only relevant aspect for him. The manager dislikes getting up very early, he hates being stuck in traffic jams and costs are not irrelevant for him either. Finally, he is reluctant to give up a cozy evening at home, which he would have to if he were to leave on Sunday.

The *uncertainties* that the manager considers to be important are: the possibility of a train delay as well as traffic conditions on the freeway, and thus the travel time. These factors cannot be predicted with certainty, but there are at least some indications that can help the manager derive *expectations* such as: "there is no rail strike announced, so I don't have to expect a delay of more than 20 minutes" or "at this time of the year, I can expect a travel time of 60-70 minutes if I drive on a Monday morning. I will need much more time, of course, if it is foggy or if a major traffic jam occurs".

The *consequences* of actions and uncertainties can be described in statements such as: "If I use the train on Monday morning and there is no delay, I will arrive at the meeting on time, spend €40 on the ticket and enjoy a trip without any stress. If I use the car on Monday morning and there is a traffic jam, it will cost me €30 and I will arrive stressed and too late at the meeting".

Within each component, a further decomposition may and should take place. Alternatives can frequently be decomposed further into sub-components. The decision of an automobile manufacturer regarding a new model can be decomposed into a large set of smaller decisions on engine size, body shape etc. Other examples of modularly composed alternatives are corporate strategies or the text of a contract. The same decomposition concepts can be applied to uncertainties by modeling them as the aggregate of many factors. As an example, the manager from the above scenario could decompose the risk of heavy traffic on the freeway into the uncertainties of fog, icy road and accident. In the same vein, the risk of a nuclear reactor accident can hardly be judged globally. One would need to make a list of all possible causes, evaluate their probabilities and derive the overall reactor risk by properly aggregating these judgments. The determination of the preference system of a decision maker always requires decomposition. The dissatisfied car owner who is considering the purchase of a new car has to split the very general objective of "having a better car" into a number of more explicit sub-objectives.

Decision analysis provides a number of techniques and tools that help the decision maker decompose a decision problem into components that are easier to handle.

1.2.5 Subjectivity

The rationality concept in decision making contains no conditions regarding expectations and preferences. To put it crudely: every decision maker can expect and like whatever he wants. Expectations and preferences are inherently subjective. They have to be justified, however, and be consistent with the principles of rationality that the decision maker has approved. In management and business decisions, for instance, "rationality" does not require profit maximization and cost minimization. Whether it is rational for a poor casual worker to stint on food in order to buy an Armani suit depends only on whether he thoroughly considered the decision and made it in accordance with his expectations and objectives.

Two people can hence make different decisions in an identical decision situation without one of them necessarily being any more or less rational than the other. The explanation of the decision mismatch can simply lie in different expectations of the future and/or different objectives.

The thorough contemplation of one's own objectives and preferences can be highly demanding for the imagination, all the more so if the consequences of a decision extend over many years into the future. In personal life, decisions concerning partner selection, family planning and career choice are good examples. In the public arena, the construction of a new airport close to a major city, the location of a storage site for nuclear waste or a new abortion law are other relevant examples. Even in private life, with respect to decisions that only affect oneself, it is often not easy to imagine what objectives and preferences one will have two, ten, or twenty years from now. If several people or even large portions of the population are involved, this task is even harder.

A tricky question is whether or not it is rational to attach more weight to present desires than to foreseeable long-term preferences. In anticipation of the lure of the sirens,[2] Ulysses asked his men to tie him to the mast of his ship and gave orders not to be untied, not even if he were to command them to do so later. Such methods of "self-commitment", i.e. a self-imposed restriction of future decision flexibility, could be rational if the decision maker foresees that his decision making in the future will be flawed, e.g. myopic (Elster 1979). An excellent example is provided by people who try to quit smoking. To avoid all temptation, they may destroy all cigarettes in their home and inform all their friends about their attempt at quitting. Another interesting example can be observed in the context of retirement savings: many people deliberately choose long-term savings products that require regular payments and are difficult to terminate once initiated. The simple reason for this choice is the anticipation of their own limited self-discipline, and this choice gives protection against myopic consumption desires that might threaten their long-term goal of making sufficient provision for old age (Normann and Langer 2002). On the other hand, it is not clear that the decision maker's present preferences are more relevant than his future interests. Many people would state the preference to die in a serious accident rather than to end up paraplegic and wheelchair-bound for the rest of their lives. After such an accident has occurred, however, they adjust to the new situation and are glad that they survived.

Expectations and objectives are not available on demand in the decision maker's head. Decision theory has developed procedures that can help develop and articulate subjective expectations and preferences. This might be a tedious process, but one can learn and become quicker. An essential benefit of devoting oneself to decision analysis is that such a commitment will over time lead to more conscious decision making that is directed by one's own expectations and objectives.

[2] Three women in Greek mythology who lured seamen with their enchanting music and voices to wreck their ships on the rocky coast of their island.

1.2.6 Incomplete knowledge and the concept of dominance

Quite often, we can articulate our expectations about uncertain events only vaguely. Everyday language is packed with fuzzy terms such as "most likely", "possibly", and "presumably". In the same fuzzy way, we express our preferences. On the one hand, we support the idea of unrestricted mobility; on the other hand, we feel that some measures should be taken against excessive traffic in the city center.

Sometimes, we are also excessively accurate in our statements. We believe that "it is completely impossible" that our long-term friend Honest could be involved in embezzlement. We would "never" consider moving to a large city (or to the countryside). But would we seriously bet everything we have on the fact that Honest will never succumb to temptation? And is there really no possible scenario that would us make consider an apartment in a large city (or a farmhouse)?

Decision analysis attempts to support decision makers in the proper articulation of expectations and preferences. However, it has to recognize that there is fuzziness. It is not sufficient to identify some alternative to be the "optimal" one, based on the stated expectations and preferences. One should also check the robustness of this result with respect to small changes and variations in the underlying parameters.

It is characteristic of modern decision analysis that it does not strive for optimization at any cost and does not overly stress the reliance on "soft" data in the process. As we will see in the forthcoming chapters, incomplete probability information and incomplete preferences may well be sufficient to substantially simplify a decision problem or even to determine an optimal alternative.

The term dominance" is used to describe a situation in which it is possible to identify some alternative *a* as being superior to some other alternative *b*, even though information about the expectations and preferences of the decision maker is incomplete. You will become acquainted with various dominance concepts while working through this text book. They differ on the one hand with respect to the type of information that is lacking (probabilities or preferences) and on the other hand with respect to the decision rule that is used to determine the ranking of alternatives. This can be illustrated by the following examples:

1. You are searching for an assistant. The characteristics you care about are his or her job-related knowledge, ambition and ability to work in a team. You evaluate each applicant on a ten-point scale with respect to each of these criteria. The decision rule you have decided to apply is to choose the applicant who earns the most points overall. However, you do not simply add up the points from each category but want to attach weights to the criteria by multiplying the points allocated to a factor before aggregating them. You realize that applicant *a* was assigned a higher point score for each and every category compared to applicant *b*. Thus, for the decision between these two applicants, you don't even have to think about weighting the factors in more detail. To make this decision, you do not need preferences that are completely defined. With respect to the decision rule "maximize weighted sum of points" alternative *b* is dominated by alternative *a*.

2. A very similar argument can be applied if you have not yet formed a clear opinion about the probabilities of the uncertain events. The profitability of a project depends on the state of the world in the future. The decision rule you decide to apply is to compare alternatives with respect to the statistical measure "expected value of profits" (i.e. the sum of profits weighted with their probability of occurrence). If the set of alternatives contains a project x that promises a higher profit than a project y for each and every state of the world, there is no doubt that it will also result in the higher expected profit. To draw this conclusion, you do not need more explicit probability information – project x dominates project y.

1.3 Applications and practical relevance of decision analysis

Modern decision analysis has mostly been developed and advanced by American scientists and consultants; hence, most applications still lie in the Anglo-American environment.

Keefer et al. (2004) provide an overview of practical applications of decision analysis methods that were published in selected journals between 1990 and 2001. This survey becomes even more insightful as it also contains a comparison with an earlier survey by Corner and Kirkwood (1991) that covers the years from 1970 to 1990. The combined information provides some insights on tendencies of development in this area. The collected applications are subdivided into the fields

- Energy
- Manufacturing and Services
- Medical
- Military
- Public Policy
- General.

With respect to the trends and developments, the authors conclude that the overall number of articles that deal with practical applications of decision analysis has increased over time. In the energy area, they observe a shift towards applications in environmental risk while the previously dominant problems of regulation and (factory) site location have become less relevant. Military applications have gained importance (they were basically non-existent in the 1991 survey and there was not even a sub-category for them). The medical area has become less prominent in the updated survey; however, as the authors argue, this finding is probably driven by the fact that decision analysis methods are so established in the medical field by now that they are published in specialized medical journals and no longer in the decision theory journals that are covered by the survey.

The latter point effectively illustrates that it is problematic to judge the dissemination of decision analysis in practice by relying on literature surveys. Many applications are performed by consultants and consulting firms and thus do not appear in the publicly accessible literature. In addition, many applications concern very important decisions and are classified as secret by the (US) government, for

example, or by certain companies. Based on these considerations, Edwards (1984) referred to the "invisibility of decision analytic practice" more than three decades ago".[3] In the meantime, the visibility of decision analysis in practice has increased considerably. This is reflected, for instance, in the fact that many highly sophisticated software tools that have been brought to the market over the last few years constitute implementations of decision analytic ideas and concepts. The "Decision Analysis Software Survey" that is compiled biennially by the specialized journal OR/MS lists, in its October 2008 issue, no less than 25 software packages that provide support in problem structuring, risk modeling, sensitivity analysis and other important aspects of decision analysis. It is furthermore interesting to note that, in the meantime, many of these products now allow joint work by an entire team and are thus suited to supporting project teams working on large-scale problems. In a similar vein, "decision conferencing", meaning an intensive computer assisted group meeting, has become increasingly important over the last few years (Keefer et al. 2004).

Nevertheless, decision analysis does not yet seem to have exploited its full potential in practical applications. One reason might be a lack of awareness. Many research insights have not advanced beyond the specialized journals and academic text books into corporate practice and applications in other institutions or private life. In addition, barriers of acceptance are likely. The decision support procedures assume that the decision maker reveals his true expectations and preferences without any reservations. This is not always in the personal interest of managers, politicians or others who have to make decisions in interaction with other people. Openness towards others limits tactical opportunities and might evoke resistance (Wrapp 1984). The transparency of the decision foundation, which is a key postulate of the theory, makes the decision maker vulnerable when the decision is monitored by others in retrospect. Finally, a manager might be afraid that using a formal procedure reduces his control over decision making. How can he decide on some procedure before knowing what decision will be produced by this method?

Even though such tactical considerations might obstruct the applicability of analytical concepts in group decisions with highly diverse opinions and preferences, the benefit to individual decision making remains unaffected.

In the published examples of practical decision analysis, the "*decision analyst*" plays a major role. The decision analyst is commonly an external consultant skilled in applying decision analytic instruments giving advice and support to the internal project team with respect to problem structuring and resolution. The decision analyst is not the decision maker but contributes the methodological knowledge, while initially lacking factual insights about the decision problem. He is unbiased and thus more suited to assuming a counseling function than any member of the decision making company, government body or other institution.

[3] If you are particularly interested in the early roots and the historical development of decision analysis, you should read Edwards et al. (2007). This collected volume also contains some interesting examples of decision analytic applications, presented in more detail.

The decision analytic way of thinking and the associated procedures are useful for decision making in many different fields. Hiring an external consultant is obviously only advisable for major decisions. The aim of this textbook is to make you sufficiently familiar with this way of thinking and the specific procedures so that you appreciate the insights and are capable of incorporating them into the future decisions that will come up over the course of your professional and personal life.

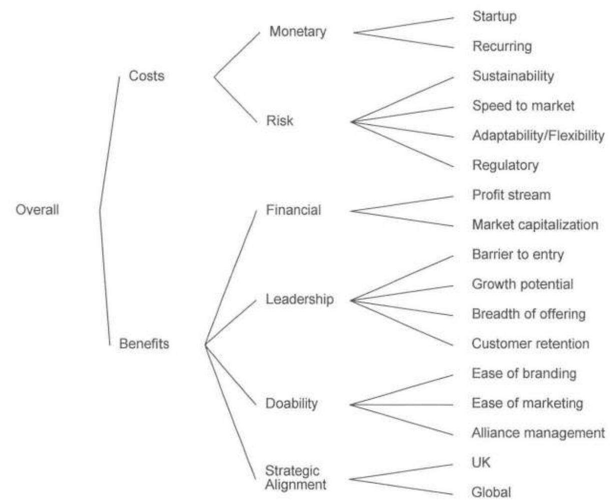
Case Study: New business appraisal

Source: Phillips (2006), pp. 382-385.

In order to strengthen the worldwide business of its group, a British financial services company considered establishing a new e-commerce activity. Initial screening of business alternatives had already generated a set of three promising options: SMB, Bank and Benefits. As none of the three alternatives dominated the others in every aspect, agreement was reached on using the approach of decision conferencing to reach a final decision.

A group meeting was called, attended by twelve company executives and four outside consultants already working on the project. After the three business alternatives had briefly been presented by the consultants, all participants were asked to privately score the options (on a metric scale) to indicate their overall preference. With ten votes for Benefits, six for SMB and zero for Bank, however, no agreement could be reached.

The group hence decided to implement a formal decision model trading off costs and benefits to obtain another ranking of the alternatives. After two hours of discussion, a decision model was developed. The structure of the model is reflected in the following figure; project risk was considered as one aspect of costs.



Subsequently, the three options were scored on each criterion by mutual agreement of the group. In case no consensus was reached, the majority view was input to the model. In contrast to the individual scoring, where most votes were cast for Benefits, SMB emerged as the clear winner.

As several participants challenged some of the weights used in the model, an extensive sensitivity analysis was conducted. Nevertheless, the superiority of SMB proved remarkably robust over large ranges of weights for the different criteria. In the end, the group agreed on SMB as the best solution for the new e-commerce venture. This stylized case highlights the main benefits of decision conferencing: a shared understanding of the decision problem and mutual commitment to the way forward within the group of decision makers is often reached.

tions for a flower garden or when developing alternatives for the formulation of bylaws.

While searching for alternatives – or generating them – the question arises of when the process should be stopped and the decision made. Sometimes, time and budget restrictions limit the further creation of alternatives. In other cases, the disadvantages of delaying the decision or additional cost of searching for further alternatives have to be weighed against the chances of finding a better solution than those determined so far. An additional complication arises if alternatives that are available right now (e.g. job offers, flats, used cars) could become unavailable when further delaying the decision.

These decisions about continuing or terminating the search for alternatives are decisions on their own. Sometimes they are trivial, compared with the actual decision problem and can be made without elaborate analyses. When looking for a used VW Golf, you can decide easily if you want to pick one from the available offers or if you would prefer to wait a week. In other cases, like combating an acute danger – an oil tanker accident, a hostage-taking, an epidemic – the choice between the available options might be less problematic than the decision to continue searching in the hope of finding a better alternative.

The decision to continue searching must be based on objectives and expectations, just like every other decision. Objectives are necessary in order to evaluate the quality of the available options and to gain an impression of which alternatives might be superior. Expectations concerning the number and quality of additional alternatives have to be formed, as well as expectations about the effort associated with the process.

Chapter 4 deals with the problem of systematically generating new alternatives and preselecting the most appropriate ones.

2.2.2 The set of alternatives

The final decision entails the selection of one alternative from a given number of options. We define the set of alternatives as A and a single alternative as a; several alternatives are a, b, c etc.

As the name implies, alternatives must be mutually exclusive. It does not make sense to decide between "going for lunch at noon" and "watching TV in the evening" when you could do both. By combining actions that do not exclude each other, you obtain a set of mutually exclusive alternatives. Assume, for example, that for lunch, there are the alternatives "going out for lunch" and "staying at home" and for the evening, there are the alternatives "watching TV" and "reading a book". This allows us to construct four alternatives that exclude each other:

- *a* going out for lunch and watching TV in the evening,
- *b* going out for lunch and reading a book in the evening,
- *c* staying home for lunch and watching TV in the evening,
- *d* staying home for lunch and reading a book in the evening.

The set of alternatives A contains at least two elements. If the number of alternatives is so large that they cannot all be checked with the same intensity – e.g. hundreds of job applications for one open position – pre-selection strategies have to be applied (see Chapter 4). One possible approach is to specify minimum requirements for the education and/or age of the applicants.

The number of alternatives is infinite for continuous decision variables. The possible amounts of money that can be spent on an advertising campaign can vary infinitely. The same holds for the production output of a detergent or the time invested by an expert in a particular project. Usually, it is possible to discretize a continuous variable without distorting the problem too much. As an example, an investor could choose a virtually infinite number of percentage values when splitting his portfolio between bonds and stocks, but he can also reduce the endless number of alternatives to the following restricted set:

- *a* 100% bonds,
- *b* 75% bonds, 25% stocks,
- *c* 50% bonds, 50% stocks,
- *d* 25% bonds, 75% stocks,
- *e* 100% stocks.

In this manner, we obtain a simplification of the decision problem, but simultaneously also a coarsening. Assume the decision maker thinks that, of the given alternatives, the allocation of 75% to bonds and 25% to shares is optimal. If he thinks the set of alternatives is too coarse, he can fine-tune in a second step by choosing from similar alternatives such as 70, 75 and 80% bonds. In this book, we will mostly focus on situations in which only a few alternatives are considered.

2.2.3 One-level and multi-level alternatives

Every single decision is a part of the universe of decisions an individual has to make. This also applies to the time dimension: you only look ahead to a certain point in time ("planning horizon"); everything beyond this point is subject to future decisions. Quite often it is foreseeable, however, what uncertainties are relevant for the future and how the decision maker could and should react to these events. No sophisticated chess player will think ahead only one step. Multi-level alternatives are also called strategies.[1] A strategy is a sequence of contingent decisions; examples of two-level strategies are:

- I will listen to the weather forecast and traffic report at 6 a.m. If both are encouraging, I will drive to work at 7:15 a.m., otherwise I will take the train at 6:59 a.m.
- An additional amount of €200,000 will be invested into the development project. If a marketable product exists by the end of the year, it will be produced. If there is no marketable product, but further development looks promising, a well-funded associate should be sought to provide financial support. If the development is not promising, it should be terminated.

[1] Usually, the term "strategy" is defined in such a way that one-level decisions are also included as a special case. Nevertheless, in everyday language, it is rather uncommon to use the term "strategy" to refer to very simple one-level decisions.

The choice of the number of decision levels that are taken into consideration is a preceding decision and similar to the issue of a further search for alternatives.

2.3 Modeling the states of the world

2.3.1 Uncertainty and probability

In a *decision under certainty*, every alternative is determined by an immediate consequence; no unknown influences affect it. In the case of *uncertainty* – synonymously, we will also speak of *risk* in the following – the outcomes depend on forces that cannot be fully controlled by the decision maker.[2] Strictly speaking, there is no decision under total certainty. Anyone could be struck by a meteorite at any time or – with a greater probability – suffer a stroke. It is a subjective preceding decision to neglect or to consider the different sources of uncertainty. Fully neglecting uncertainty simplifies the problem in general, because only one state of the world has to be dealt with.

The reason why the uncertainty can be neglected is usually not that it is insignificant; however, it is not necessary to account for uncertainty in the calculations if it is foreseeable that one alternative will turn out to be optimal for all scenarios. The optimal solution is independent of uncertain events, for instance, if the decision can easily be reversed. If a decision is irreversible or can only be reversed at large cost, the risk has to be considered in the calculation. In many situations, uncertainty is the key problem; this is often the case for large investments, medical treatments, court decisions and political decisions of the legislature.

If you decide to consider uncertainty, it has to be formalized in a model which includes one or several uncertain facts (also known as chance occurrences). An uncertain fact is a set of *outcomes*, of which exactly one will occur. The set of outcomes is exhaustive and mutually exclusive, e.g. the outcome of a soccer game between the teams A and B can be described by the set $\{A$ wins, B wins, A and B draw$\}$.

The outcomes will occur either as *events* or as *states*. The result of a football game and the resignation of a CEO can be regarded as events, whereas the presence of crude oil in a certain geological formation or the health status of a patient can be regarded as states. The distinction between states and events is irrelevant from a formal perspective.

[2] A word of warning might be appropriate here. The use of terms in the literature is not very consistent in the domain of risk and uncertainty. Often, uncertainty is used as a general term to subsume "risk" (the existence of a definite probability distribution of the consequences) and "ignorance" (knowledge of the set of possible consequences, but not at all about their probabilities). We depart from these definitions because we think that the concept of ignorance is evasive and theoretically dubious (see also Section 10.1). Nevertheless, in the case of risk, the beliefs of the decision maker concerning the probability distribution can be more or less incomplete. The literature often uses the term "ambiguity" to describe this type of vague probability information (see also the discussion on ambiguity aversion in Section 13).

Event and state variables are intrinsically discrete or continuous. The number of marriages on a specific day in a specific registry office, for instance, would be a discrete variable. The amount of rainfall, on the contrary, would be a continuous variable: in principle, there is an infinite number of possible amounts of rainfall for each single recording. However, how to treat an uncertain fact in the modeling of a decision situation is a question of expedience. In the same spirit as in our earlier discussion about decision variables, in many cases it makes sense to discretize a continuous chance variable. For example, for a very large investment decision it might be sufficient to vary the uncertain acquisition costs only in millions of dollars.

Table 2-1: Some uncertain states and events

Uncertainties	States or events
What will the weather be like tomorrow?	dry rainy
What will the dollar exchange rate in Frankfurt/Main be on Dec. 1, 2010?	€0.65 €0.70 €0.75 €0.80
Is Patient X infected with tuberculosis?	Yes No
How will the union react to the employers' wage raise offer?	Accept Decline, but willing to negotiate Decline, strike ballot

Without loss of generality, we start by assuming a finite set of states. To each state s_i a *probability* $p(s_i)$ is assigned. In order to qualify as a probability, the figures $p(s_i)$ have to fulfill the following three conditions (Kolmogoroff 1933):

- $p(s_i) \geq 0$ for all i.
- $\sum p(s_i) = 1$ (the certain state has the probability 1).
- $p(s_i \text{ or } s_j) = p(s_i) + p(s_j)$ (the probability of occurrence of one of several disjoint states equals the sum of the probabilities of the states).

2.3.2 Combined events or states (scenarios)

A decision situation is often best described by the combination of several uncertain facts. For example, in a decision problem, the quantity of US sales of a specific product, as well as the dollar exchange rate, might be relevant. The setting is therefore described appropriately by combinations of sales figures and exchange rates.

Events or states that are composed of various uncertainties are called "data constellations" or "scenarios". In the case of complex decisions, a great number of uncertainties have to be considered, they may stem, for instance, from the technical, economic, legal, social or political environment. In such cases, the modeling of

Chapter 2: Structuring the decision problem

2.0 Summary

1. The basic structure of a decision problem entails alternatives, uncertainties, consequences of alternatives and uncertainties, as well as the objectives and preferences of the decision maker.
2. Important decisions on the relevant alternatives have to be made in advance. They refer to such questions as: Should further alternatives be considered or should a choice be made from the existing ones? Should the number of existing alternatives be reduced by merging similar alternatives or increased by splitting existing alternatives into several variants? Should the options be designed as one-stage or multi-stage alternatives?
3. Other important preceding decisions relate to the modeling of uncertainties. Can the future be predicted sufficiently well to neglect uncertainties in general? If not, what are the relevant uncertainties that influence the outcomes of the decision problem? In how much detail or how general should the universe of possible states be modeled?
4. Uncertainty is described by states or events to which probabilities are allocated. Probabilities have to obey certain rules: joint probabilities and conditional probabilities are relevant to combining uncertain events. These combinations can be visualized by the use of event trees or cause trees.
5. If an alternative is chosen and uncertainty resolved, a certain consequence will be obtained in a deterministic fashion. It might be necessary to formulate an effect model that determines the consequences.
6. When modeling the preferences, a preliminary decision has to be made on whether a single objective or several objectives should be considered. The relevant objectives have to be identified. In the case of uncertain expectations, risk attitude has to be considered, and in case of consequences that occur at different points in time, it might be reasonable to model time preference.
7. Usually, it is not possible to model the different components of the decision problem independently of one another. The components influence each other. The decision maker moves back and forth between the set of alternatives, the uncertainty structure and his model of preferences until he finishes modeling the decision problem and derives the optimal decision.
8. Graphical forms of representation such as the influence diagram, the decision matrix, and the decision tree are very useful tools. They force the decision maker to clarify his conceptions and support him in communicating with and explaining the decision fundamentals to other people involved in the decision process.

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_2,

2.1 The basic structure

The basic assumption of prescriptive decision theory is that a complex decision problem can be solved more effectively by decomposing it into several components (separate aspects). Instead of dealing with the problem as a whole, the decision maker analyzes the components and creates models of the problem's components. Afterwards, the partial models are merged to generate an overall model of the decision situation. These components were already mentioned in Chapter 1:

1. The *alternatives* (synonymous: options, actions). The decision maker has a number of alternatives from which to choose;
2. The *uncertainties*. These are incidents or states of the world that have an influence on the decision, but cannot be controlled at all or at least only partially by the decision maker. The decision maker can only form expectations about the resolution of uncertainty;
3. The *consequences* of actions and uncertainties. By choosing an alternative and the resolution of uncertainty, the resulting consequence is determined. This does not necessarily mean that the result is immediately *known*. An "effect model" might be needed to specify which consequences follow from the decision variables and event variables;
4. The *objectives and preferences* of the decision maker. The decision maker has different preferences with respect to the consequences, i.e. he usually prefers one outcome over another. If no objective that the decision maker considers relevant is affected by the decision, there is no serious decision problem to solve.

Modeling is by no means unique; the same problem situation can be depicted in multiple ways. The remainder of this chapter covers several aspects and tools of modeling. The later chapters go into more detail on several key components.

2.2 The modeling of alternatives

2.2.1 The problem of finding alternatives

In some cases, finding the relevant alternatives is no problem; they are given in a "natural" way. The manager who learns at 6 a.m. about the closure of the highway because of fog can choose between the trains at 6:59 a.m. and 7:13 a.m.; there are no other alternatives if he wants to be on time for his meeting. The jury can declare the defendant to be guilty or not guilty. The voter can mark one of the given alternatives in the voting booth or return a blank or invalid sheet.

In many other situations, acceptable alternatives are not known immediately; generating them may be a considerable part of the problem. This can be a *search process*, as for instance that for someone who wants to buy a used car in a metropolitan area. It can also be a creative process of *generating* alternatives, e.g. when looking at different ways of constructing a machine, when reflecting on design opuncertainty is also called "scenario analysis". The number of relevant event or state sets can become overwhelmingly numerous. For four uncertain facts with three states each, there are already $3^4 = 81$ scenarios to be considered. The probability of each scenario cannot be determined by simply multiplying the probabilities of the states. This is possible only in the most often unrealistic case of independent events (see Section 2.3.3). The effort of defining and calculating probabilities for the scenarios has to be weighed against their usefulness. In particular, in the context of strategic planning, scenarios form the basis of incorporating uncertainty. It is desirable to have just a small number of distinct scenarios that can be used for evaluating risky strategic alternatives. A practical example that deals with the definition of such scenarios to model the world energy needs can be found in the case at the end of this chapter.

The combination of events can occur multiplicatively or additively. The first case is given if two or more events are intended to occur jointly. The second case is given if we are looking for the probability that out of multiple events, precisely one will occur. Consider the example of the two uncertain facts "What will the weather be like tomorrow?" and "What will my mother-in-law do tomorrow?". Each of these uncertain facts is described by an event set consisting of two events that are relevant to your decision:

- What will the weather be like tomorrow? = {dry, rain}
- Will my mother-in-law visit us tomorrow? = {visit, not visit}.

You might be interested in the probability of the scenario "tomorrow the weather will be dry and your mother in law will not visit". The probability of the scenario for both events happening jointly is derived by multiplying the probabilities of "dry" and "will not come".

On the other hand, you may be interested in the probability of the scenario "it will rain tomorrow *or* your mother-in-law will visit". To fulfill the scenario, it is sufficient here for one of the two events to occur. In such a case, the two relevant probabilities have to be combined in an additive manner.

In the following sections, both cases are discussed in more detail.

2.3.3 The multiplication rule

The concepts of *conditional probabilities* and *joint probabilities* are relevant to the conjunction of uncertain states. Let x be an event from the set of events X, and y be an event from the set of events Y. The *conditional probability* $p(y|x)$ is then the probability that y will occur, given that x has already occurred. This conditional probability is defined for $p(x) > 0$ as

$$p(y|x) = p(x,y) / p(x). \tag{2.1}$$

In this definition, $p(x,y)$ refers to the *joint probability* of x and y. This is the probability that x and y will both occur.

From (2.1), it follows that

$$p(x,y) = p(x) \cdot p(y|x). \tag{2.2}$$

Eq. (2.2) is known as the multiplication rule and can be used to calculate the probability of the joint occurrence of two events. Let us look at an example in order to explain these concepts. The uncertain fact X indicates the rate of economic growth of a country for the next three years. It is necessary to differentiate between three states: (x_1) depression, (x_2) stagnation, (x_3) boom. Y stands for the results of the next general election; in this case, we only distinguish between the events (y_1) victory of the conservatives and (y_2) victory of the socialists. Assume that the economic forecasts for the next years look like this:

$$p(x_1) = 0.2 \quad p(x_2) = 0.65 \quad p(x_3) = 0.15.$$

The probabilities for the election outcome depend on the rate of economic growth. It is assumed that the lower the rate of growth, the better the socialists' chances. The following conditional probabilities are formed:

$$p(y_1|x_1) = 0.4 \quad p(y_2|x_1) = 0.6$$
$$p(y_1|x_2) = 0.5 \quad p(y_2|x_2) = 0.5$$
$$p(y_1|x_3) = 0.6 \quad p(y_2|x_3) = 0.4.$$

Using (2-2), this information allows us to calculate the joint probabilities which are listed in Table 2-2. For example, the combination (scenario) "boom and socialist victory" has a probability of $p(x_3,y_2) = p(x_3) \cdot p(y_2|x_3) = 0.15 \cdot 0.4 = 0.06$.

Table 2-2: Joint probabilities of economic growth and political development

			Y (political development)	
		$p(x_i)$	y_1 (conservative)	y_2 (socialist)
X (economic growth)	x_1 (depression)	0.20	0.08	0.12
	x_2 (stagnation)	0.65	0.325	0.325
	x_3 (boom)	0.15	0.09	0.06
	sum	1	0.495	0.505

By summing over the joint probabilities in each row or column, the unconditional probabilities $p(x)$ and $p(y)$ of the two sets of events can be determined. They are also referred to as *marginal probabilities*. The marginal probabilities for economic growth were given; those for political developments are derived from the joint probabilities in the table; for instance, the conservatives will win with a probability of $p(y_1) = 0.495$.

The calculation of joint probabilities can also be represented graphically. In Figure 2-1, each circle reflects a set of events that consists of several alternative events, depicted by the branches originating from the knots. The numbers are the probabilities assigned to these different possibilities of economic growth. Further to the right, the numbers reflect the conditional probabilities of the election result depending on the economic growth and the resulting probabilities for the event combinations of economic and political development.

In the given example, the conditional probabilities $p(y|x)$ differ, i.e. the probability of a conservative or socialist victory in the elections is dependent on the rate of economic growth. A simpler situation would be given if all the conditional probabilities were the same; in this case, the unconditional probabilities would equal the marginal probability and the probability of a specific election outcome would in fact not be dependent on the economic development.

Independence
Two events x and y are referred to as (stochastically) *independent* if for each y_j the conditional probabilities $p(y_j|x_i)$ are identical for each i and thus

$$p(y|x)=p(y). \tag{2.3}$$

Inserting this equation into (2.1), we obtain

$$p(x,y)=p(x)\cdot p(y), \tag{2.4}$$

i.e. the joint probability of two independent events equals the product of their marginal probabilities.

2.3.4 Event trees

Event trees can be useful tools for depicting scenarios. An event tree starts with an uncertain fact that may lead to one of several possible events; each of these events can be followed by further events. The leaves of the tree (the triangles on the right) reflect event combinations (scenarios) that are mutually exclusive. Their probabilities can be calculated by multiplying the probabilities along the respective path. With the exception of the probabilities at the root of the tree (i.e. the uncertain fact on the left), all the probabilities are conditional. Depending on the context, the expression state tree might be more appropriate than that of event tree.

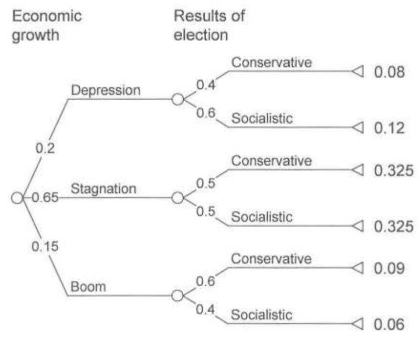


Figure 2-1: Event combinations and their probabilities derived by the multiplication rule

The graphical representation of the election scenario in Figure 2-1 is a very simple example of an event tree. As a further example, we consider the event tree that was used in the "Reactor Safety Study" of the US Nuclear Regulatory Commission from 1975 (the so-called Rasmussen report). It was generated to analyze the probabilities of serious reactor accidents. As can be seen in Figure 2-2, a major pipe bursting in the primary system is used as the starting event (the event "pipe does not burst" is not shown); subsequently, the reaction of other parts of the system is considered. More explicitly, the internal power supply, the emergency cooling system, the disposal system for nuclear fission waste and the container system are modeled in detail. For each of these system components, only two events – the system component works or fails – are considered.

Every path through the event tree stands for the theoretical possibility of a sequence of accidents. Not all the sequences are logically meaningful. If, for example, the electricity supply fails, none of the other system components can work. An event that occurs with probability zero and all the following events can be eliminated from the tree. Accordingly, the event tree can be simplified as shown in the lower part of Figure 2-2.

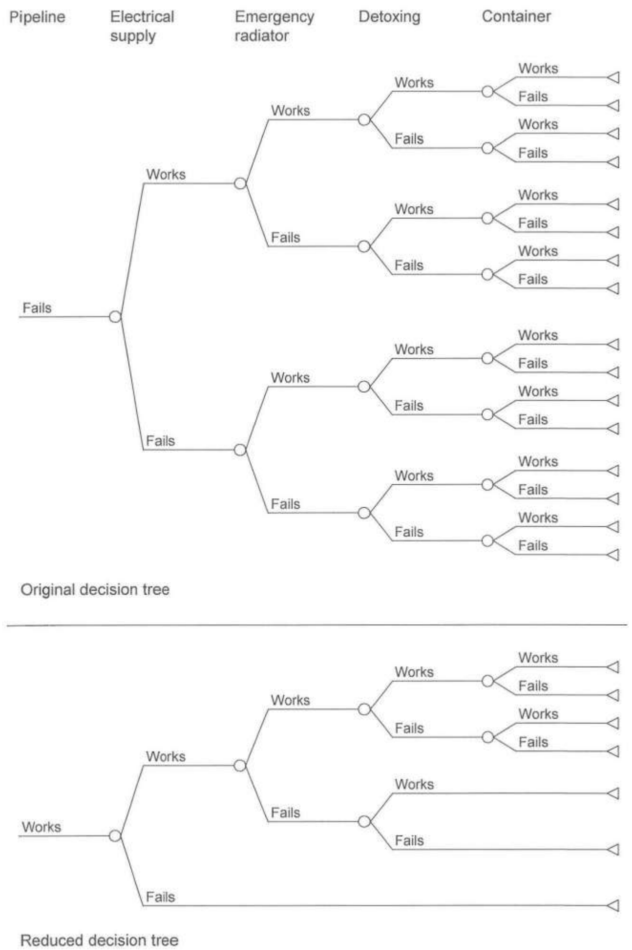


Figure 2-2: Simplified event trees for a major reactor accident resulting from a burst pipe in the cooling system. Source: Bunn (1984), p. 171.

2.3.5 The addition rule

The probability of either x or y or both occurring is

$$p(x \text{ or } y) = p(x) + p(y) - p(x,y). \tag{2.5}$$

The subtractive term becomes clear if we realize that both $p(x)$ and $p(y)$ already include $p(x,y)$; this term is therefore counted twice and has to be subtracted.

Assume, for instance, that a farmer estimates the probability of pest infestation to be $p(x) = 0.2$ and the probability of drought to be $p(y) = 0.15$. How high is the probability that the crop is destroyed if each of the two catastrophes can cause total destruction? In order to make this calculation, we need the probability of the subtractive term $p(x,y)$, i.e. the probability of pest infestation and drought occurring simultaneously. If infestation and drought are stochastically independent, this value is $0.2 \cdot 0.15 = 0.03$. The danger of crop loss is then $0.2 + 0.15 - 0.03 = 0.32$. However, it is also possible that there is stochastic dependence, such as a higher probability of infestation in the case of drought, compared with periods of humidity. Assuming the farmer estimates the (conditional) probability of infestation in the case of drought to be 1/3, we then obtain $p(x,y) = 0.15 \cdot 1/3 = 0.05$ and the danger of losing the crop is $0.2 + 0.15 - 0.05 = 0.3$.

If x and y are mutually exclusive, i.e. they cannot happen at the same time, the subtractive term equals zero. If the pest does not survive a drought and the probabilities $p(x)$ and $p(y)$ stand as stated above, the probability of losing the crop is $0.2 + 0.15 = 0.35$.

2.3.6 The cause tree

A second instrument that played a major role in the preparation of the earlier-mentioned reactor safety study is the *fault tree*, which reverses the idea of the event tree. The starting point is a predefined final result and the aim is to determine how it could or did happen. As the name suggests, this method was constructed for analyzing accidents – reactor accidents, plane crashes, failures of automobiles. Fault trees are suited to analyzing the causes of malfunctions of complex systems. Like an event tree, a fault tree is not restricted to unpleasant or negative events, which is why we prefer to use the more neutral term "cause tree".

A cause tree starts with the effect and attempts to determine the possible causes. For each cause, it then continues to consider what could have induced it.

In event trees, we only observe multiplicative compositions ("and"-conjunctions) between events, i.e. the joint occurrence of several events. In cause trees, however, we also find "or"-conjunctions, which require the addition of probabilities.

The mode of operation of a cause tree is depicted in the example in Figure 2-3. The Manifold corporation owns 20% of the shares of Simplex corporation, which competes in the same markets. The management of Manifold is interested in a majority holding. As the Simplex shares are widely spread, but not traded on an exchange, the management is considering a public takeover bid and offering the Simplex shareholders an attractive price for their shares.

Two problems have to be taken into account. There is the possibility that the activity will not produce the desired majority in voting rights, and in the event that it does, there is still the possibility that the cartel office will not allow the takeover because of violations of cartel law.

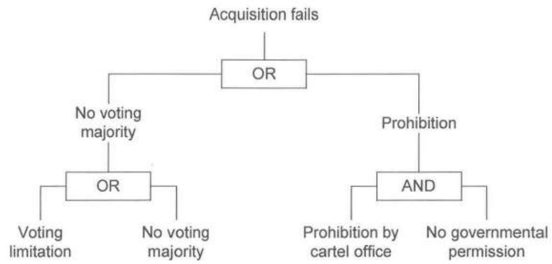


Figure 2-3: Cause tree for possible failure in a takeover

Simplex's bylaws include a restriction on voting rights, stating that no single shareholder can exercise more than 5% of the voting rights. If this clause is not rescinded in the next general meeting, owning the majority of shares does not help Manifold. It has to be feared that the Simplex management will counteract the takeover and the change in bylaws will not be pushed through. In addition, Manifold may not even obtain the majority of shares.

The other potential obstacle is the cartel office. It could prohibit the takeover, because it sees the danger of a market-dominating position in some sectors. In the case of a prohibition, there is still the possibility that the commerce secretary will issue special permission, overruling the cartel office.

What is the probability of Manifold's endeavor failing under these circumstances? "Failing" can be caused by the events "no majority in voting rights" or "prohibition by the cartel office". These two results have to be linked with an "or". According to (2.5), it holds that

$$\begin{gathered} p(\text{failing}) = \\ p(\text{no majority in voting rights}) + p(\text{prohibition}) \\ - p(\text{no majority in voting rights, prohibition}). \end{gathered}$$

However, the cartel office will only start an investigation and possibly intervene if a majority of voting rights for Manifold seems likely. Therefore, the joint probability $p(\text{no majority in voting rights, prohibition})$ equals 0 and it holds that:

$$\begin{gathered} p(\text{failing}) = \\ p(\text{no majority in voting rights}) + p(\text{prohibition}). \end{gathered}$$

For the event "no majority in voting rights", there are again two possible causes linked by an "or": the voting cap and the possibility of not attaining 50% of the

shares. The probability of an overall failure because of a failure to obtain a majority in voting rights may be decomposed as follows:

$$\begin{gathered} p(\text{no majority in voting rights}) = \\ p(\text{voting cap}) + p(\text{no majority of shares}) \\ - p(\text{voting cap, no majority of shares}). \end{gathered}$$

We will next look at the probability of the prohibition. For this to happen, the cartel office has to forbid the takeover *and* the commerce secretary has to refuse to issue special permission. The probability of this scenario equals the product of the probabilities of the cartel office forbidding the takeover and the (conditional) probability that, given a prohibition by the cartel office, the secretary will not overrule the prohibition and will deny a special permission:

$$\begin{gathered} p(\text{prohibition}) = \\ p(\text{cartel office forbids}) \cdot p(\text{secretary does not overrule} \mid \text{cartel office forbids}). \end{gathered}$$

2.3.7 The dependence of the uncertainty model on the objectives

Unquestionably, there is an endless number of ways to model the uncertain environment. The decision as to which of the uncertain facts the decision maker takes into account and with which sets of states he chooses to model the uncertainty should be primarily influenced by his objectives.

Consider two men who are both thinking of buying a certain piece of land. Their decision is influenced by uncertainty. One of the potential buyers is a farmer who wants to grow tomatoes on the land. For him, the success of the decision depends on whether competitors will settle in the same area, whether import relief for foreign tomatoes can be expected and whether the use of insecticides will be restricted by law. He defines scenarios as combinations of different levels of competition, trade regulations and environmental laws and tries to determine probabilities of these scenarios arising.

The other potential buyer plans to build a fun park on the land and to exploit it commercially. He is obviously not interested in any of the uncertainties the farmer worries about; instead, he cares about other uncertainties like the population development in the area, the costs of construction and maintenance and possible subsidies by the municipality. The scenarios he considers are totally different from those of the farmer.

This example highlights the fact that objectives play a key role in the modeling process. We discussed this aspect before when looking at the process of generating new alternatives. Not only the compilation of the set of alternatives, but also the choice of relevant events and states, must be guided by objectives. If the decision maker does not (yet) know what he wants to achieve, he cannot identify the relevant uncertain states.

2.4 The modeling of consequences

If one alternative was chosen and specific states realized (i.e. the relevant uncertainties have been resolved), we assume the occurrence of a unique consequence. If the decision is made to choose the alternative "I will not take an umbrella along when I go for a walk" and the weather assumes the state "thunderstorm", then the consequence "I will get wet" is certain.

The consequence is not always that easy to determine. It might be necessary to use an impact model that uniquely defines the consequence of the decision. A trading company's earnings before taxes from an export deal P_T are given by a function of the selling price s, the quantity of sales (depending on p) q, the purchase price k, the tax rate t and the exchange rate x. The decision alternatives are the different selling prices. The other variables are uncertain figures. The equation

$$P_T = (1-t) \cdot q(s) \cdot (p \cdot x - k)$$

is the model that defines the consequence P_T by combining the decision variable (s) and the state variables (q, k, t, x).

The model can consist of one equation or a system of equations, but it can also be a complicated algorithm. Take as an example a production division in which many customer orders wait to be processed. For each order, what is known is which machines are needed, how long the order will take, in what sequence the order occupies the machines and for which date the delivery is planned. In the case where several orders wait in line for a specific machine, i.e. more than one order is ready to be processed by a machine, a priority rule can be used. Examples of such rules are: "first in, first out", "shortest time occupying the machine is processed first" or "closest delivery deadline is processed first". The decision problem is which priority rule to establish. As soon as a new rule is introduced, a totally different sequence plan is scheduled. This does not mean that the plan is already known; it has to be determined first. In order to accomplish this goal, a special algorithm is used that determines, usually software-supported, when to start which order, when to process it with which machine and when which order is completed. Only by the application of this algorithm does the impact of the chosen alternative on the objective variable become known.

2.5 The modeling of preferences

2.5.1 Objectives and preferences

"Preferences" are the decision maker's attitudes towards alternatives with respect to their consequences. We distinguish between the following relations for $a, b \in A$:

$a \succ b$	a is preferred to b.
$a \sim b$	indifference between a and b.
$a \succeq b$	a is preferred to b or there is indifference.

The preferences regarding the alternatives are not given beforehand; the decision maker usually has no coherent perception of them. It is the aim of decision analysis to support the decision maker in deriving them. To arrive at this point, it is important to have a clear understanding of the decision maker's preferences regarding the consequences that result from states or events of relevant uncertainties and from choosing a particular alternative.

In a first step, the decision maker has to figure out which aspects of the consequences have an impact on his preferences and are therefore relevant for him when solving the decision problem. Before buying a car, he could realize that reliability and costs are two aspects that affect his preference. However, the car's overall eco-balance, the sum of environmental effects during production, usage and disposal, does not matter to him. Combined with a statement about the direction of his preference, the decision maker hereby identifies his *objectives*. In this specific case, they would be "minimization of costs" and "maximization of reliability". Furthermore, the decision maker has to think about which specific *characteristics* he wants to use to describe explicitly the relevant consequences of his decision. These characteristics are also called *attributes*, *objective variables* or *target variables*. In the case of buying a car, he could describe the reliability in terms of the frequency of breakdowns, which he wants to minimize. The cost objective could be operationalized by a suitable combination of purchase price and running expenses. In the above-mentioned problem regarding the sequential processing of orders, attributes, such as mean pass-through time, number of missed deadlines or mean machine utilization could be chosen.

The level of valuation of an objective variable often decreases or increases monotonically with the value of the objective variable: lower costs and higher reliability are always better. In other cases, the optimal values are somewhere in the middle of the range. A person on vacation is interested in warm weather, but not the maximum possible heat; a surfer needs wind but not a hurricane. In these cases, we can say the objectives are "the most pleasant temperature" or "the best wind for surfing".

2.5.2 Conflict of objectives

One of the precedent decisions is to specify by how many and by which characteristics the consequences are defined. Chapter 3 will deal with this problem. In many economic decisions, one can concentrate on a *single* objective variable, like earnings or costs. It is a typical feature of "hard decisions" that *various* objectives exist that conflict with each other. The conflict is that there is no alternative that is better – or at least not worse – with respect to any objective variable; you cannot have everything. Solving the conflict always requires making trade-offs. The transition from alternative a to alternative b might cause an improvement for some objectives but a deterioration for other objectives at the same time.

In the following discussion, we will assume in most cases that the total value of a consequence for a decision maker results from the simple aggregation of the evaluations for the various relevant characteristics. We know this principle from many practical applications such as product tests, sports (e.g. decathlons) or analytical performance evaluations. In all these cases, the importance of the different aspects is accounted for by using weights or point schemes. In decision analysis, the use of an additive aggregation model is quite common. In contrast to the mostly naïve application of such concepts in practice, decision analysis clearly prescribes under what circumstances an additive model is acceptable and how to proceed, so as to obtain evaluations that are consistent with the decision maker's preferences. We will discuss these questions in more detail in Chapters 5 and 6.

2.5.3 Risk preferences

In the case of *decisions under uncertainty*, the decision maker's attitude towards risk plays an important role. In decisions under certainty, the problem is restricted to choosing between (certain) consequences – a problem that can be challenging enough if there are conflicting objectives. The best consequence determines the best alternative. In the case of uncertainty, one has to choose between alternatives that can lead to different consequences. Each alternative is represented as a bundle of possible consequences, each occurring with some probability. In the literature, it is common to speak of choosing between "lotteries". Given the choice between buying a block of flats from one's savings, an investment which has a low but certain return, and engaging in some speculative investment transactions, e.g. buying some high-risk securities, a risk-averse investor might choose the first while a risk-seeking investor prefers the latter. Neither of them could be labeled as acting irrationally, however. Decision analysis explains how to measure subjective risk attitude and how to make complex decisions under uncertainty considering this individual risk attitude. We will deal with these issues in Chapters 9 and 10.

2.5.4 Time preferences

The implications of a decision are often spread out over a considerable period of time, one could say that a decision has several consequences distributed over time. Individuals are usually not indifferent between the temporal spreading of consequences over time – instead, they have *time preferences*. People tend to postpone unpleasant surgery, for example, but they prefer to go on a cruise this year rather than next year. The decision about the correct point in time to start the retirement savings process is a typical question of time preference. It is a trade-off between consuming today and the perspective of having consumption opportunities in old age. In order to evaluate consequences that are spread widely over time, it is necessary to model time preferences. This problem will be discussed in Chapter 11.

2.5.5 Modeling preferences by functions

In modern prescriptive decision theory, decision makers' preferences are modeled by functions. These functions assign evaluations to the consequences or outcomes, in order to reflect the preference. In the case of certain expectations, the preference functions are called *value functions*; under risk, they are called *utility functions*. These functions are derived from preference statements in very simple choice problems – or at least from problems that are much easier to solve than the decision problem we are interested in. If the decision maker is able to give consistent answers, a value function or a utility function can be derived. This derivation is based on axioms that are commonly accepted as principles of rational behavior. The function can then be used to evaluate more complex alternatives.

In contrast to many criteria that are suggested in theory and practice and are more or less arbitrarily-defined decision rules, the procedures and concepts of decision analysis have the advantages that

- they aim to model the "true" preferences of the decision maker and that
- the evaluation of alternatives is founded axiomatically, i.e. if the decision maker accepts a few basic rationality postulates, the evaluation and optimal decision follows logically and unequivocally.

2.6 Recursive modeling

The basic principle of decomposing a complex problem into modules that can be handled more easily separately does not mean that these modules are independent of one another. It is almost never possible to model alternatives, uncertainties and objectives completely separately from one another. We have already pointed out the superordinate function of objectives several times in the last few sections.

Figure 2-4 symbolizes how these components influence one another. Because of these influences, it is not possible to generate the sub-models in a single linear run. Instead, a change in one of the sub-models can also cause the need for a revision of another sub-model. The decision maker thus goes back and forth repeatedly between the different sub-models in order to adjust them to one another optimally. We call this process recursive modeling and will illustrate this procedure by means of the following example.

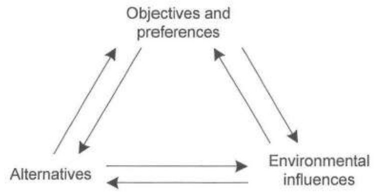


Figure 2-4: Mutual impact of the sub-models

Let us assume you are considering purchasing a notebook to replace the desktop PC that you have been using at home so far. You talk to a friend who owns a "JapTop" notebook and is very satisfied with it. After this conversation, your decision situation can be described as follows:

1. Alternatives:
   - I buy the "JapTop", or
   - I gain a better overview of the market situation first or
   - I decide not to purchase a notebook for now.
2. Uncertainties:
   - How quickly will a model that I buy today be outdated and unable to work with modern software?
   - Will low-budget models frequently cause problems that require money and time to resolve?
   - To what extent will I need the notebook in the near future for work that I could not simply do at home at my desk?
3. Objectives:
   - The notebook should be as low cost as possible.
   - It should be as powerful as possible.
   - It should be as easy to handle as possible.
   - I would also like to be able to use the notebook on campus, in particular, when researching literature in the library, so I do not need to transcribe my scribbling later at home.

After this first modeling of the situation, you turn to computer shops for advice and study relevant magazines.

Alternatives → Objectives: You come to know more and more models, and since these have different characteristics which spark your interest, you develop new objectives. For instance, you realize that there are pleasant and less pleasant keyboards and that especially weight, display size and battery longevity vary considerably. Thus, the enhancement of the set of alternatives causes an enlargement of the system of objectives.

Objectives → Alternatives: However, the inverse effect occurs as well. The longer your wish list becomes, the higher your motivation to search for better alternatives. New opportunities occur; others are dismissed – the "JapTop", which you almost bought in the beginning, might not even be a serious option any longer. In addition, you enlarge your set of alternatives by not only taking classic notebook models into consideration. Instead, you now also think about buying a subnotebook that could be used in combination with the desktop at home.

Uncertainties → Objectives: You think about how quickly the notebook will be outdated and will have problems with modern, memory-intensive applications. You realize that it is very important for you that you can still use the notebook for your final thesis in two years. This objective, which will probably eliminate the option of buying an older model, was not part of your explicit set of objectives before.

Objectives → Uncertainties: It is one of your objectives to keep costs as low as possible. Therefore, you make an effort to collect information on how prices of notebooks will probably develop during the upcoming year and whether larger

price jumps can be expected. It also matters to you what technological improvements can be expected in the near future.

Alternatives → Uncertainties: By chance, you run into a bargain offer. A notebook that has hardly been used and is in a performance class far above the one you were originally interested in is offered to you at a discount of 40% compared to its original price. It still costs more than you intended to spend. The bargain offer would only pay off if you also needed the notebook for modern computer games with special graphics requirements. Due to your challenging field of study, it seems questionable whether you will have any time at all during the upcoming years to play computer games.

Uncertainties → Alternatives: Contemplating the intensity of notebook usage that can be expected, you realize that you often need to print out (multicolored) slides for some lectures in order to make notes during class. You realize that if you bought a tablet PC instead of a regular notebook, you could insert your comments directly into the electronic slides and not only save considerable printing costs, but also help save the environment. You now also take this alternative into consideration seriously.

At some point in time, you have to push yourself to make a decision. The modeling of alternative actions, uncertain facts, objectives and preferences has to be terminated at some point. Based on the resulting model of the decision problem, you make a decision. This decision could also be to refrain from buying a notebook at all for the time being.

2.7 Visualization of decision situations under uncertainty

2.7.1 Benefits of graphical representations

Structuring and modeling a decision problem aims to support the decision maker in better understanding the problem and increasing the rationality of the solution. The means of representation we will discuss in the following support this goal. They force the decision maker to be clear and precise in phrasing objectives, alternatives, influences and consequences. In addition, they allow the decision maker to convey his perspective of the problem to other people in a clearer and less ambiguous way than would be possible with a purely verbal description.

The three types of graphical representation that we will discuss below – the influence diagram, decision matrix, and the decision tree – play different roles in the decision process. The influence diagram aims to provide a comprehensive impression of the general structure of the decision problem. It provides an overview of the interaction of the problem constituents (decision components, uncertainties, objectives) that are considered to be relevant. Such an overview is important to understanding whether sub-problems can be separated and addressed in isolation, and at what stage of the decision process what information has to be available. To ensure clarity, most details are kept out of an influence diagram. In particular, it is not explicitly displayed which alternatives are under consideration and what specific uncertainty scenarios are regarded. To deal with such details, decision matrices and decision trees will be employed in a later stage of the decision process. The two types of graphical representation are very similar with respect to their informational content (and we will discuss this issue further in Section 2.7.5). They both display full information about alternatives, uncertain events and consequences (and thereby of course also information about the objectives of the decision maker). In particular, the decision matrix arranges and presents the relevant data in a way that most easily allows us subsequently to derive a numerical solution.

2.7.2 The influence diagram

Influence diagrams (Howard and Matheson 1984, Oliver and Smith 1990) play an important role in the problem structuring phase, i.e. in an early stage of the decision analysis process. The great relevance of this tool can be illustrated by the fact that the special issue "Graph-based presentations" of the scientific journal Decision Analysis (Horvitz 2005) contains mostly articles that deal with influence diagrams. The survey article by Howard and Matheson (2005) that appeared in this special issue is by far the most frequently cited article that ever appeared in Decision Analysis.

Influence diagrams do not display all possible actions but only the decisions per se. The set of alternatives is represented by a single symbol (rectangle) that does not convey how many and which alternatives exist. Likewise, not each single event but only the overall set of events is represented by a circle or an oval. Additionally, the single consequences do not appear but only the objective variables, symbolized by diamonds or hexagons.

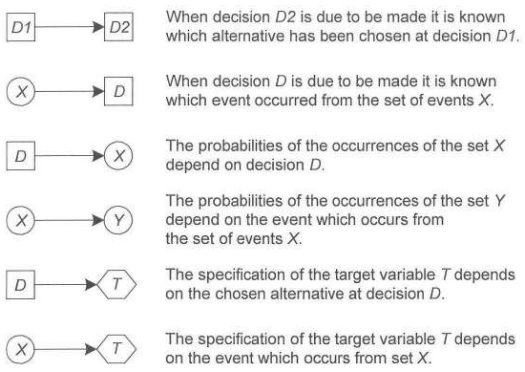


Figure 2-5: The presentation of relations in an influence diagram

Arrows pointing to a decision symbol depict a piece of information available at the time of the decision. Arrows pointing to an event symbol mean that the event's probabilities depend on the directly preceding event or the directly preceding decision.

If two event symbols are connected by an arrow, this indicates stochastic dependence between them, but does not necessarily have to indicate causality. In principle, the direction of an arrow could just as well be reversed, because if X is stochastically dependent on Y then also Y is stochastically dependent on X. If an arrow is missing, the events are independent of each other.

Figure 2-5 summarizes the most important constellations. Cycles are not permitted, i.e. there must not be a path through the diagram with identical starting and end points.

Let us begin with a simple case: a manufacturer of car accessories has developed a new anti-theft device. The question arises of how large a production capacity should be chosen. In order to forecast the sales potential, the decision has been made to offer the product in a local market for a few months. The price has already been determined and the costs are known. Depending on the sales in the test market, the probability distribution for the countrywide demand can be predicted and a decision about the production capacity can be made. Profit is the only objective variable to be considered. Figure 2-6 shows a suitable influence diagram for this problem.

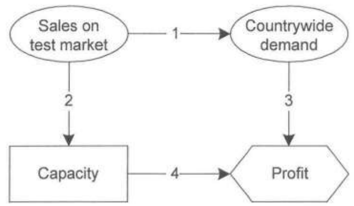


Figure 2-6: Influence diagram for the manufacturer's decision concerning production capacity

Let us take a closer look and begin with the sales volume in the test market, which is still uncertain at the present time. When this figure becomes known, the countrywide demand can be assessed (arrow 1). This can refer to both a deterministic forecast and a probability distribution for the countrywide demand. Arrow 2 indicates that sales in the test market are known before the decision on the capacity is made. Arrows 3 and 4 depict that the profit is influenced by both the countrywide demand and the chosen production capacity.

In this example, you can see that the direction of the arrow between the two sets of events "Sales in test market" and "Countrywide demand" could also be reversed, which would represent the true causality. However, from the decision maker's point of view, the chosen representation is the natural one as it is consistent with the chronological order: first, sales in the test market are known, then the assessment of countrywide demand results.

Let us consider a somewhat more complicated case that is based on a study by Jensen et al. (1989). It deals with the decision of a US state whether or not to require the use of smoke detectors in residential buildings by law. Furthermore, if this decision is made in favor of the detectors, a decision has to be made on the degree of the state's effort to enforce this regulation. The extent to which a regulation by law is able to reduce the number of casualties and injuries depends on numerous factors. The more home owners voluntarily install smoke detectors anyway, the smaller the effect. In addition, a certain refusal rate has to be taken into consideration; not everybody will adhere to the law. This can be influenced by the intensity of enforcement of the regulation, e.g. via inspections. Further influence factors obviously are the fire frequency and the failure rate of smoke detectors. In addition to the reduction in the number of human casualties, further goals of the measures are the reduction in financial damage and the minimization of private and public costs. The different influences are represented in Figure 2-7.

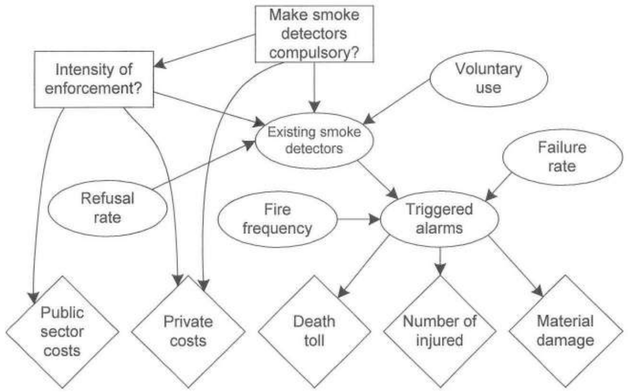


Figure 2-7: Influence diagram for the decision on whether to regulate the use of smoke detectors

One of the strengths of influence diagrams is their assistance in structuring a problem. A second strength is the ability to communicate and document the relevant decisions and uncertain influences in a well-arranged manner. Of course, the identification of the influences is not sufficient to make a decision; in fact, in a second step, these influences have to be quantified in a model. In the previous example, for instance, it would be necessary to estimate on a statistical basis how the number of human casualties and the magnitude of financial damage is related to the number of fire alarms triggered.

This detailed information is intentionally not integrated into the influence diagram, however; otherwise, the diagram could no longer serve its purpose of providing a good overview of the general structure of the decision problem. Nevertheless, even despite the lack of detail, influence diagrams can become very extensive for complex decision problems. An example is shown in Figure 2-8, depicting influence factors on possible health effects 10,000 years after the closure of a nuclear waste disposal site. This excerpt of an influence diagram displays which uncertain factors might influence the objective variable in combination with the construction of the barrier system. Even though this presentation is no longer particularly clear, it is definitely better suited to formulating and documenting views of complex interactions in a collaboration of experts than is purely verbal explanations.

Afterwards, the decision alternatives as well as the relevant uncertainties and consequences have to be determined in order to initiate the concrete steps of problem solving; possibly, this happens only for isolated sub-problems). The decision matrix and decision trees discussed below are suitable forms of illustration for the relevant facts of the decision problem.

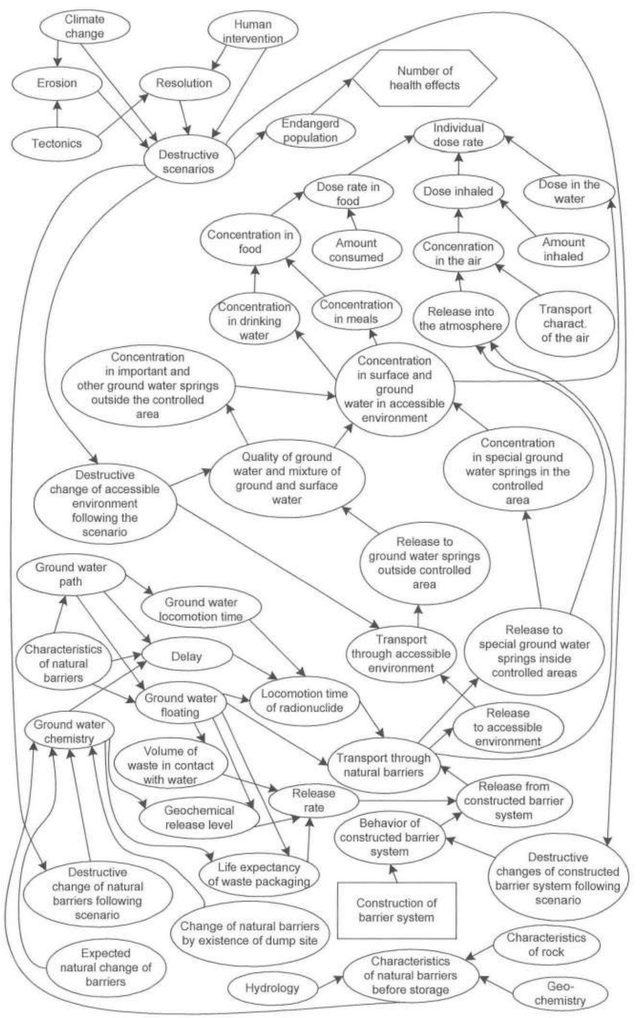


Figure 2-8: Influence diagram (excerpt) to evaluate nuclear waste disposal sites (Merkhofer 1990).

2.7.3 The decision matrix

Let A be the finite set of alternative actions and let S be the finite set of possible and mutually exclusive events. We assume that by pairing any alternative $a \in A$ and any state $s \in S$, a resulting consequence c_{as} is uniquely determined. If each row of a matrix represents an alternative and each column an event, then, each cell may be used to display a respective result (consequence). If there is only one objective, each consequence is described by the value that the objective variable assumes. For multiple objectives, it is represented by the vector of parameter values for all objective variables; this is illustrated in Table 2-3. In the left matrix, a_i stands for the assumed value of the objective variable of alternative a given that state s_i occurs. In the right matrix, a_{ij} refers to the value of the jth objective variable if alternative a is chosen and state s_i occurs which happens with probability $p(s_i)$.

Table 2-3: Decision matrices with one and multiple objective variables

	s_1	…	s_i	…	s_n
	$p(s_1)$	…	$p(s_i)$	…	$p(s_n)$
a	a_1	…	a_i	…	a_n
b	b_1	…	b_i	…	b_n
c	c_1	…	c_i	…	c_n

	s_1	…	s_i	…	s_n
	$p(s_1)$	…	$p(s_i)$	…	$p(s_n)$
a	$a_{11},\dots,a_{1m}$	…	$a_{i1},\dots,a_{im}$	…	$a_{n1},\dots,a_{nm}$
b	$b_{11},\dots,b_{1m}$	…	$b_{i1},\dots,b_{im}$	…	$b_{n1},\dots,b_{nm}$
c	$c_{11},\dots,c_{1m}$	…	$c_{i1},\dots,c_{im}$	…	$c_{n1},\dots,c_{nm}$

Let us illustrate the case with only one objective by means of the following example. Think of a publisher who wonders how many copies of a book he should produce and stock. He considers 5,000, 7,000 or 9,000 printed copies as the relevant alternatives. The uncertain environment is described by the demand occurring at the given price. The publisher considers the states 4,000, 5,000, 6,000, 7,000, 8,000 or 9,000 demanded books to be possible. The only relevant objective variable is the profit.

The market price will be €15; fixed costs for the entire process with any alternative are given as €10,000. The proportional costs per copy amount to €10. The profits charted in Table 2-4 result from the model

$$P = \min(C,D) \cdot 15 - 10 \cdot C - 10{,}000,$$

where C refers to the number of copies and D to the demand.

Table 2-4: Publisher's decision matrix

	Demand					
Number of copies	4,000 (0.10)	5,000 (0.15)	6,000 (0.15)	7,000 (0.30)	8,000 (0.20)	9,000 (0.10)
5,000	0	15,000	15,000	15,000	15,000	15,000
7,000	–20,000	–5,000	10,000	25,000	25,000	25,000
9,000	–40,000	–25,000	–10,000	5,000	20,000	35,000

A rational solution to this decision problem requires the publisher to think about the probabilities that he assigns to all possible levels of demand. For instance, if in all likelihood the demand will not be higher than 6,000 copies, a small batch of books should be produced, e.g. 5,000. However, if the expected demand can be assumed to 8,000 or 9,000 copies, a much higher supply seems reasonable.

In the example, the probabilities that the publisher assigns to the different demand levels are charted in Table 2-4 (numbers in brackets). Since the set of states in a decision matrix needs to be comprehensive and the states need to be mutually exclusive, the sum of the probabilities is one.

If multiple objectives are of importance, the values of all objective variables have to be inserted into the cells. Let us assume that the publisher is not only interested in profits, but also wants to avoid disappointed customers (who do not receive a copy because demand exceeds supply). Consequently, he considers the number of customers who cannot be served as a second objective variable. We then obtain the following decision matrix in Table 2-5.

Table 2-5: Decision matrix of the publisher with two objective variables, profit (€) and disappointed customers (D)

	Demand					
Number of copies	4,000 (0.10)	5,000 (0.15)	6,000 (0.15)	7,000 (0.30)	8,000 (0.20)	9,000 (0.10)
5,000	0€ 0 D	€15,000 0 D	€15,000 1,000 D	€15,000 2,000 D	€15,000 3,000 D	€15,000 4,000 D
7,000	–€20,000 0 D	–€5,000 0 D	€10,000 0 D	€25,000 0 D	€25,000 1,000 D	€25,000 2,000 D
9,000	–€40,000 0 D	–€25,000 0 D	–€10,000 0 D	€5,000 0 D	€20,000 0 D	€35,000 0 D

2.7.4 The decision tree

For the visual representation of multi-stage alternatives, the decision tree is often better suited than the decision matrix. A decision tree contains the following elements:

- decisions, represented by squares,
- uncertainties, represented by circles or ovals,
- consequences, represented by triangles.

Lines representing alternative actions emanate from each decision square; lines representing alternative events or states emanate from each uncertainty circle. At every event symbol, the sum of the probabilities has to equal one. Each path across the tree from left to right ends in a consequence.

Figure 2-9 shows an example. A company needs to decide whether to continue or to abort the development of a new product. The probability of successfully completing the development is 0.3. If successful, the company needs to decide whether to develop large or small production capacities. The probability of high demand for the newly developed product is 0.6 while the probability of a low demand is 0.4.

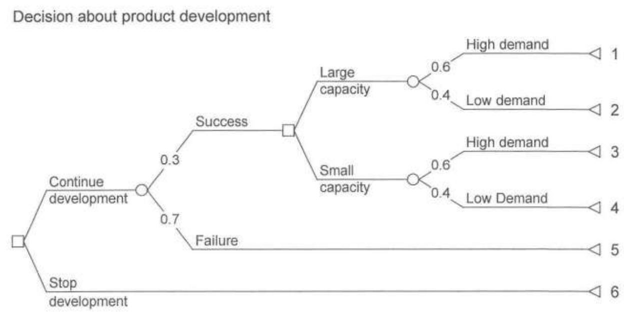


Figure 2-9: Decision tree for the product development problem

Due to a lack of space, the consequences are only labeled with numbers. For an exact description, the respective values of the objective variables have to be given. If the objective was purely financial, for instance, consequence 2 could follow from the development costs, investment costs for constructing the large production capacity and the marginal returns from the sales in the case of low demand.

Representing a decision situation in a decision tree usually provides some design flexibility. On the one hand, complex *alternatives* can be *split up* into subsequent actions. For example, a company whose space capacities do not suffice anymore could think about either expanding the building at hand, purchasing some ground and constructing a new building, or purchasing an already completed building. For each of these alternatives, two variants can be distinguished. They are represented in a "two-stage" model in part (a) of Figure 2-10. Obviously, however, the alternative actions can equivalently be presented as six "one-stage" alternatives, as can be seen in part (b) of the figure.

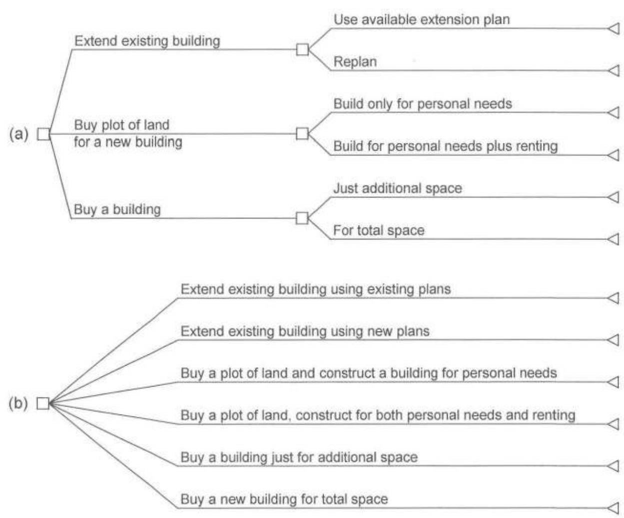


Figure 2-10: Equivalent representation of alternatives

On the other hand, *events* can be combined or split up. In a situation where a company faces the risk of running out of raw material, because of an impending strike at a supplier's production site, it might be advisable to evaluate the general probabilities of a strike (and its length) first. In a second step, the conditional probabilities of material shortage are assessed for both a short and a long strike. Figure 2-10 contains two equivalent representations (a) and (b) for this case. In part (b), the probabilities result from multiplying the probabilities for the different durations of a strike with the conditional probabilities for the possible material supply consequences in part (a).

If, for the problem at hand, the strike itself is irrelevant and only the material supply matters, this uncertain influence can be reduced to the two events "material shortage" and "no shortage". These probabilities result from adding the respective mutually exclusive cases from part (b) and are represented in part (c) of Figure 2-11.

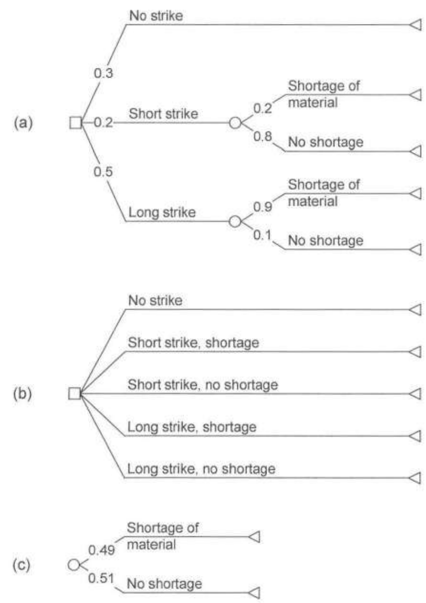


Figure 2-11: Equivalent representation of events

All the strategies a decision maker has at hand can be read off the decision tree. To describe a strategy, the decision maker has to specify for each decision that could occur which alternative he would choose if he were to reach this point in the decision process. To depict a strategy in a decision tree, you would thus need to mark at each square, one (and only one) of the lines extending to the right. Returning to the example from Figure 2-9, the strategy "continue development and (if successful) choose small capacity" could thus be depicted as shown in Figure 2-12. Overall, this procedure would produce four different combinations of arrows (2×2). Obviously, however, we can condense two of these strategies, because, when choosing "abort development" in the first decision stage, the capacity decision in the second stage will be purely hypothetical, as it cannot be achieved in the tree anymore. In the example of product development, there are thus three strategies to consider:

a. Continue development. If successful, provide large capacity,
b. Continue development. If successful, provide small capacity,
c. Abort development.

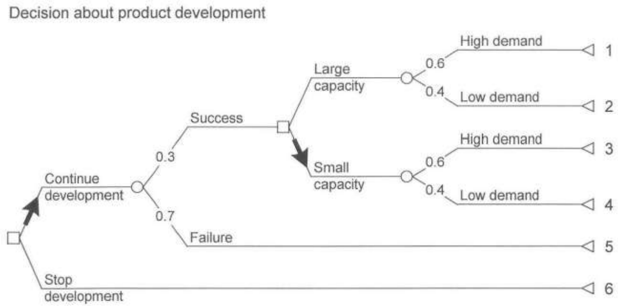


Figure 2-12: Representation of a strategy in a decision tree

Likewise, the scenarios can be derived and depicted in the decision tree. Scenarios can be seen – in a manner of speaking – as the "strategies of the environment". In our example, there are again four possible strategies with two of them condensable.[3] The following three scenarios remain:

1. Development successful, high demand,
2. Development successful, low demand,
3. Development unsuccessful.

Figure 2-13 depicts scenario 2.

[3] Strictly speaking, the three uncertainty knots, each with two possible events, produce eight different combinations (2 × 2 × 2). However, chance cannot select different paths for the two knots "demand" (at least not if we assume that the chosen capacity does not influence the demand – probably a safe assumption). Therefore, the number of sensible combinations is reduced to four.

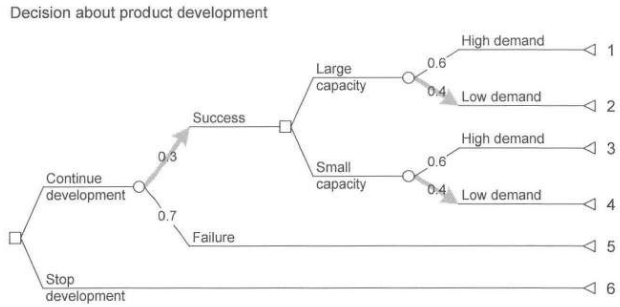


Figure 2-13: Representation of a scenario in a decision tree

This type of representation can at the same time effectively illustrate how the encounter of a strategy and a scenario results in a unique consequence. In Figure 2-14, we combined the strategy *b* from Figure 2-12 with the scenario 2 from Figure 2-13. As a result, we obtain a unique path through the complete tree. If strategy *b* is chosen and chance "plays scenario 2", we obtain consequence 4.

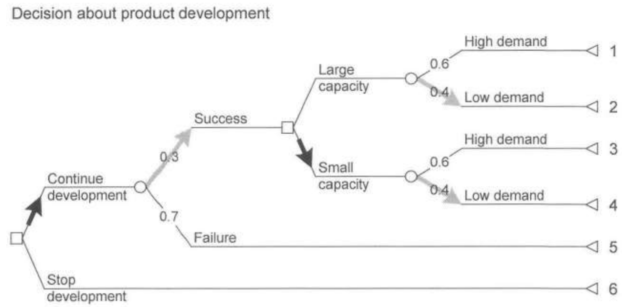


Figure 2-14: Strategy and scenario in a decision tree

2.7.5 Connection between decision matrix and decision tree

A decision matrix can always be transformed into a decision tree. It then consists of a single decision knot. At each alternative branch there is one event knot with all possible states. Conversely, any decision tree can be transformed into a matrix representation. This is achieved by contrasting strategies (rows) with scenarios (columns) in a table and entering the resulting consequences into the table cells (see Figure 2-14). Table 2-6 shows the decision matrix for the example.

Table 2-6: Decision matrix for the product development problem

	Development successful, high demand	Development successful, low demand	Development not successful
	$p=0.18$	$p=0.12$	$p=0.7$
a continue development. If successful, large capacity	Consequence 1	Consequence 2	Consequence 5
b continue development. If successful, small capacity	Consequence 3	Consequence 4	Consequence 5
c abort development	Consequence 6	Consequence 6	Consequence 6

This bilateral transformability shows that the decision matrix and decision tree essentially contain the same information. Even though, for a multi-stage problem, you would not obtain the original tree if you first collapsed it into a matrix and then transformed it back to a tree as described above (it would no longer be a multi-stage tree). This would, however, not matter for the determination of the optimal solution.

Questions and exercises

2.1
You want to give a birthday present to your sister. On your shopping tour, looking for the appropriate present, you find the following things that she would like:

- A teddy bear (€15),
- A book, *Daughters of Horror* (€12.80),
- A bottle of champagne (€32).

At this point, you decide to stop searching. You do not want to spend more than €50 and also consider giving her a new €50 note wrapped in wrapping paper instead of one of the items listed above. How many alternatives do you have?

2.2
You read a newspaper article about an unemployed citizen of Lazyville who has won the state lottery. You have hopes that the person mentioned in the article is your unsuccessful cousin Peter who lives in Lazyville. Thirty percent of the population of Lazyville are foreigners and the unemployment rate is 8%. Fifteen percent of the foreigners are unemployed. What is the probability of the winner being a native (i.e. not a foreigner)?

2.3
You think about wearing your new leather jacket on the way to the gym as you would like to show it to your friend who you might meet there. Unfortunately, a number of valuables have been stolen recently. There is a possibility that your jacket might be stolen while you are training.

(a) Which scenarios are relevant to this decision problem?
(b) You estimate the probability of meeting your friend at the gym to be 60% and the probability of your jacket being stolen to be 10%. What are the probabilities of the scenarios identified in (a)?

2.4
There are two events, x and y. Given the joint probabilities $p(x,y)=0.12$; $p(x,\neg y)=0.29$ and the conditional probability $p(y|\neg x)=0.90$ ($\neg$ indicates the complementary event).

(a) Calculate the probabilities $p(\neg x,\neg y)$, $p(\neg x,y)$, $p(x)$, $p(y)$, $p(\neg x)$, $p(\neg y)$, $p(x|y)$, $p(y|x)$, and $p(x|\neg y)$.
(b) Calculate the probability $p(x \text{ or } y)$, i.e. the probability of at least one event occurring?

2.5
Your brother in law Calle Noni runs an Italian Restaurant. Recently, he has been complaining about his decreasing profits. As you are studying business administration, he asks you for advice. You do not know very much about his restaurant and plan to visit Calle to gather as much information as possible. For preparation, draw a cause tree containing all possible reasons for a decrease in profits.

2.6
On Friday morning, the owner of a restaurant thinks about how many cakes he should order for Sunday. In the event that the national team reaches the finals, he expects only a few guests and the sale of only two cakes. If the national team loses the semi-final on Friday afternoon, he expects to sell 20. The purchase price per cake is €10 and the selling price is €30. The owner wants to choose between the alternatives "two cakes", "five cakes", and "ten cakes". His objective is to maximize his profit. Generate a decision matrix.

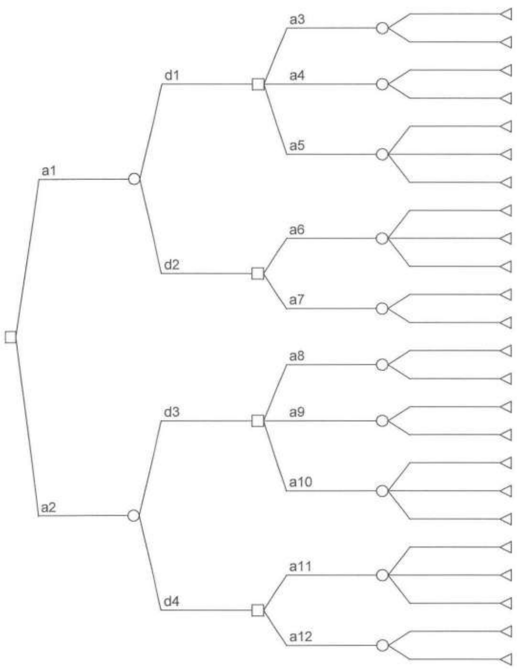
2.7
You want to go shopping and think about whether you should take an umbrella. If it rains and you do not have an umbrella, you will have to take your clothes to the

dry cleaner. On the other hand, you hate carrying an umbrella and often leave it behind at a shop. As the weather forecast will be on the radio soon, you think about postponing the decision.

(a) Structure the problem by drawing a decision tree. Indicate alternatives, events, and consequences.
(b) Is it also possible to depict the problem in the form of a decision matrix?

2.8

(a) How many strategies are in the following decision tree?
(b) Pick one of them and mark its possible consequences.
(c) How many scenarios are included in this decision tree?
(d) Indicate one of the scenarios by marking all events which happen in this scenario.



2.9

You think about donating a fraction of your million Euro inheritance to create a sports and leisure center. Clearly, the economic success of such a center depends on many factors. Depict them in an influence diagram.

Case Study 1: Bidding for the "Kuniang"

Source: Bell (1984), pp. 17-23.

An American public utility holding company, the New England Electric System (NEES), was pondering whether to bid in the auction of the "Kuniang" a ship that had ran aground off the coast of Florida in 1981. The ship could be used to haul coal from Virginia to its coal-powered stations in New England. However, a law that had passed in 1920 restricted American coastwise trade solely to vessels built, owned and operated by Americans; the "Kuniang", however, was a British ship. Another law from 1852 provided a way out: it permitted a foreign-built ship to be regarded as American-built if the previous owners declared the ship a total loss (which the owners of the "Kuniang" had done) and the cost of repairs was at least three times the salvage value of the ship.

The cost of repairing the "Kuniang" was estimated at around $15 million. However, it was unknown on what basis the US Coast Guard, who was the responsible authority, would determine the salvage value. The scrap value of the "Kuniang" was clearly less than $5 million. If the Coast Guard, however, considered the amount of the winning bid as an indication of the ship's true salvage value and NEES chose to bid more than $5 million, they would have to find some way to increase the cost of repairs accordingly. One way to increase the cost of repairs was to install self-unloading equipment in the ship. This would cost an extra $21 million and shorten the round trip voyage from 8 days to 5, which would be beneficial. However, the presence of the equipment would also lower the storage capacity, which would be unfavorable. It was uncertain how the Coast Guard would decide; NEES attached a probability of 70% to the bid price and 30% to the scrap value.

Since the transport capacity of the "Kuniang" exceeded the needs of NEES by far, the profitability of the acquisition was highly dependent on the freight rates that could be charged for additional transports from external clients.

As for the chances to be awarded the contract, NEES was sure that a bid of $10 million would win and that a bid of $3 million definitely would not. For the values in between, the following probabilities were assumed:

	$5m	$6m	$7m	$8m	$9m
Probability	1/6	2/6	3/6	4/6	5/6

As alternatives to the "Kuniang", with or without the self-unloading equipment, two offers from American shipbuilders regarding new ships were available.

The situation is depicted in the following decision tree. For simplification, only the offers of $5, $7, and $9 million were plotted and uncertainties regarding the freight rates were not depicted. The numbers marking the consequences stand for the expected net present values of the investment. Since one of the offers for building a new ship proved to be superior, the other one was not considered anymore.

Improvement, High-Tech-Future and New Society) were created out of the aforementioned existing energy scenarios. For instance, the scenario "Business-as-Usual" assumed a moderate GDP growth and a higher environmental impact. This scenario extrapolated current trends in energy growth and productivity, whereby a decreasing use of nuclear power and hence an increase in the use of fossil fuels and a lower promotion of alternative energies were assumed. Additionally, another scenario (Hard Times) was created in order to integrate unexpected negative developments into the analysis. This scenario was effectively not supposed to reflect actual beliefs about future developments but to serve as a kind of stress test.

Such a scenario analysis can be very helpful in a political decision-making process, for example to develop hedging strategies for potential threats of negative environmental impacts or to enable policymakers to better judge the consequences of their decisions (e.g. concerning the promotion of alternative energies).

Chapter 3: Generating objectives and hierarchies

3.0 Summary

1. Rational decision making requires a clear understanding of the underlying objectives. Comprehension of objectives also helps improve recognizing better alternatives and attractive decision opportunities.
2. Attempting to gain clarity with respect to personal objectives can be mentally demanding. However, there are a number of indicators that can be focused on and harnessed: shortcomings in the status quo, comparison of alternatives, strategic goals, external guidelines and the objectives of other people that are involved.
3. Fundamental objectives have to be distinguished from means objectives. Only objectives that are fundamental in the given decision context should be considered for an evaluation of the alternatives.
4. The system of (fundamental) objectives which are relevant in a given decision situation should fulfill a number of requirements: completeness, lack of redundancies, measurability, independence and conciseness.
5. The hierarchical structuring of the (fundamental) objectives supports the fulfillment of the requirements mentioned in point 4.
6. For each (fundamental) objective used in the evaluation of the alternatives, an objective variable (= attribute) has to be determined.
7. Unfortunately, artificial and proxy attributes have to be used instead of natural attributes quite often.
8. Attributes should have the following desirable properties: direct, operational, understandable, comprehensive and unambiguous.

3.1 The relevance of objectives

Decisions are made in order to achieve objectives. We judge different alternatives (of taking actions) by how close they bring us to our objectives. Without a clear understanding of the objectives, it is impossible to make a sensible choice between different alternatives.

However, having a thorough knowledge of the objectives is also a prerequisite for *finding or generating new, thus far unknown or unrecognized, alternatives*. A careful definition of the objectives may lead to the insight that the available set of possible actions does not provide a satisfactory solution and may help suggest where to look for better alternatives. Additionally, without knowledge of the objectives, it is impossible to identify the relevant environmental influences, since the word "relevant" implies that there need be some impact on achieving the objectives. Taking this one step further, raising the awareness of one's own long-term objectives helps to better recognize one's decision opportunities. This chap-

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_3, © Springer-Verlag Berlin Heidelberg 2010

ter owes much to the book "Value-Focused Thinking" by Ralph Keeney (1992). One of the key messages of that book is that a systematic and focused way of thinking will allow you to seek out or generate decision situations which are useful in pursuing your objectives. A person who is aware of his or her own objectives keeps an eye out for decision *opportunities*, rather than just responding to decision *problems*. How often do people just do the things they have always done, that their environment expects them to do, or that turn out to be the most obvious alternative? In doing so, they thereby forego the chance of making decisions that bring them closer to their own objectives.

3.2 The generation of objectives

Objectives do not simply *exist*. Instead, they have to be developed and modeled by a thorough thinking through of the relevant issues. This can be done in peace and quiet, though often, stimuli from outsiders can be of great help in realizing which goals are of importance (Bond et al. 2008). However, people generally tend to avoid making such efforts. This can be rational in particular cases e.g. if further thinking or discussion is unlikely to generate further objectives that could result in different choices. Certainly, before making momentous and hard decisions it is always sensible to work thoroughly on defining your goals and objectives (Keeney 2007); this requires a systematic approach and some routine.

The key issue is the sources a decision maker can draw on in a specific decision context if she seeks clarity about her objectives.

1. *Shortcomings* of the status quo or expected future states of the world. Suppose, for example, that a shop owner receives more and more customer complaints concerning the service quality provided by his shop assistants. This generates the objective of improving customer satisfaction. Alternatively, suppose a doctor gives his patient a warning of a heart attack due to increased blood fat levels. The patient then becomes aware of the objective of avoiding atherosclerosis.
2. *Comparison of available alternatives.* What are the differences between the available alternatives and which of these differences are important to you? Suppose, for example, that you have to hire a secretary and wish to assess the different applicants. They will differ in their foreign language competence, computer skills, business education, age, accent, appearance, and so forth. Which of these differences are important to you and which are not? If a particular difference is important to you, this indicates a relevant objective.
3. *Strategic goals.* Some objectives are not directly linked to a certain decision situation but are of overall interest. Corporations pursue strategic goals such as technological leadership, reputation, sustainable production, or enhanced motivation of their employees. Individuals may pursue strategic objectives such as conservation of health, well-being of their family, perfecting a foreign language, or improving their social standing.

If strategic objectives are well-defined and established (e.g. as part of a company's mission statement), people can use them in specific situations. Deciding on the packaging of a certain product, for instance, offers the opportunity to pursue strategic goals such as reputation and environmental protection. When deciding on a new job offer, it is helpful to consider strategic objectives like conservation of health and the well-being of your family.

4. *External guidelines*. Within organizations, subordinate entities have to follow the guidelines of superordinate divisions. Usually, these guidelines include quantities of production and sales, turnover, costs, capital expenditure and profit ratios. The existence of such guidelines should indicate which objectives are considered important from an overall perspective and, thus, could be relevant in a concrete decision situation.
5. *The people involved*. Those who decide on the objectives should ask themselves who will be affected by the consequences of the decision and what objectives these persons potentially have. In planning your personal career, for example, the first step is to be clear about your own objectives. In a second step, though, you should take into account what objectives your family may have and how relevant they are to you in your career choice. This also applies to corporate decision making, in which the decision maker should have a clear understanding of the preferences and interests of bosses, colleagues and subordinates. Top management, in turn, always has to balance the preferences of shareholders, creditors, consumers and employees. The need to take into account third party interests becomes particularly obvious in decisions made by (public) authorities, the government or parliament.

3.3 Fundamental objectives and means objectives

3.3.1 Elimination of means-ends relations

In the context of this book, it is important to distinguish between fundamental objectives and means objectives. A fundamental objective is an objective that is pursued for its own sake and that needs no further justification.

We will put this definition into perspective in the next subsection: a fundamental objective is *fundamental only in a given context*. A means objective, by contrast, is pursued because it helps achieve other more fundamental objectives.

In marketing planning, for example, a maximum of brand awareness is aimed for, since this is considered as a *means* of maximizing profits. Another example is the minimizing of the emission of greenhouse gases which is considered to be important for preventing global warming.

Often, the management literature on business objectives (Barrett 2006) does not distinguish between fundamental objectives and means objectives; however, from a prescriptive point of view, this distinction is essential.

In order to distinguish between fundamental objectives and means objectives, you always have to ask yourself the question: "*why* is this objective important?" If an objective X is important only because it contributes to achieving another objective Y, you should reconsider whether it makes more sense to drop the means objective X from the list and replace it by the fundamental objective Y instead. There are some considerations in favor of this:

1. *More alternatives*. Identifying Y as a fundamental objective may draw your attention to new alternatives that might be suitable for accomplishing the fundamental objective. Suppose that the municipal council is thinking about how to enforce speed limit adherence on a particular road, because a number of accidents have already occurred there. Among other alternatives, traffic humps and radar speed checks are discussed as possibilities. However, speed limit adherence is not the fundamental objective, as the primary goal is the prevention of accidents. Reflection on this fundamental objective possibly raises the council's awareness of new alternatives: installing a pedestrian light or employing (school) crossing guards would be two further options.
2. *Double count*. "Double counting" occurs if a system of objectives includes a fundamental objective Y and a (with respect to Y) means objective X. This may cause erroneous decisions. In practice and in the literature on production management, "processing time minimization", for example, is an important objective. But this objective can hardly be considered an end in itself; a reduction in processing time is intended to cause a reduction in the cost of working capital (apart from that, in some cases, this may also improve the market position, as shorter delivery times can be offered to customers, but we will ignore this aspect). Suppose that a company considers two different investment alternatives for cost reduction: increase in capacity or improved quality control in the production process. With a capacity increase, order processing time could be reduced, which in turn would decrease the stock of work in progress and hence would reduce the cost of working capital. Improved quality control during the production process, in turn, would save costly error handling of finished goods. For both alternatives, investment costs have to be weighed against potential savings. It may be possible that improved quality control leads to higher savings than does increased capacity, but the management (wrongly) nonetheless opts for the latter, because this alternative results in an apparently additional and important reduction in processing time.
3. *Uncertain impact*. It remains uncertain whether and how strongly X is instrumental in achieving Y. In other words, is X an appropriate means of achieving Y at all? Some cities, for instance, obviously pursue the objective of banishing cars from the inner city by making car drivers' lives a misery. Amongst others, road bottlenecks and a lack of public parking are well suited to achieving this end. Presumably, though, it is not a fundamental objective of the city council to bother car drivers per se; rather, it is a means

objective to improve air quality and reduce noise pollution. However, it remains unclear to what extent the means is able to contribute to the achievement of the fundamental objective. It is quite possible that pursuing the means objective of "banishing car drivers from the city" ultimately causes even more traffic jams, pollution by emissions, and noise than before.

4. *Ambiguous effects*. The means objective X may be conducive to achieving Y, but may be detrimental to achieving another fundamental objective Z at the same time; therefore, the extent to which pursuing X is indeed congruent with your preferences is ambiguous. Improved product quality is definitely instrumental in terms of sales opportunities and market share, but it increases production costs at the same time. If an increased profit is aimed for, maximizing the quality of products is not necessarily an appropriate objective.

A well-designed system of objectives should not include any means-ends relations: i.e. the system should not contain objectives that gain relevance only by having an assumed impact on other objectives in the same system of objectives. Means-ends relations contain *factual judgments*, i.e. they make a statement on assumed interdependencies. They are important factors for developing impact models and for generating possible sets of alternatives. However, they have nothing to do with objectives, which make a statement about what the decision maker wishes/desires and thus include *value judgments*. Marketing managers can form expectations about whether their product will sell better as a result of greater brand awareness. Regarding the traffic policy, politicians can build some expectations on whether the stressed car driver will behave as intended, because of the measures taken. Professors and politicians may speculate about whether study conditions will be improved by reducing the duration of study. In many cases, only experts are able to evaluate the instrumentality of means objectives – but even experts are not always in a position to make a definitive statement. By contrast, identifying the fundamental objectives is not an expert's business but has to be carried out by a person who decides on her own or is in a position to make decisions for others.

3.3.2 Context-dependence of fundamental objectives

The distinction between fundamental and means objectives is relative. In a given context, an objective is defined as being fundamental if it is not a means to achieve another objective in the same context. Consider the means-ends chain $X \rightarrow Y \rightarrow Z$. In this case, X is a means objective of the more fundamental objective Y, which in turn is a means objective for the most fundamental objective Z. If Z is irrelevant for a given decision, Y then becomes a fundamental objective.

It seems almost impossible to go back to the most fundamental objectives for every little decision we make. The more fundamental the objective, the more universal the set of alternatives that has to be taken into account. Suppose you are interested in three special package tour offers from a travel agency. One of the objectives that come to your mind is "as much sun as possible." But why is sun important to you? For one thing, you can get a suntan. And why is a suntan important to you? Possibly due to increased self-esteem. Suppose increasing your self-esteem is one of the fundamental objectives that you identify. Should you take this objective into account when evaluating the package tours? If the answer is yes, you should also consider other options than the three special offers. What means – other than getting a suntan – are there for increasing one's self-esteem? For instance, writing an article that is subsequently published, or improving your service at tennis would be two of many possible further alternatives. You thus not only have to choose among the three package tours but would have to evaluate a much bigger set of alternatives, including writing an article in a highly ranked journal or practicing your serve for three weeks. Assuming you have enough time and desire to solve this extended problem… then is up to you! Without a doubt, however, it is impossible to optimize the rest of your life with respect to every single decision you make.

For this reason, it is often necessary to use "fundamental objectives" in a given decision situation that would be means objectives in an extended context. For the marketing planner, maximizing brand awareness may be of great importance when designing a new advertising strategy. Accordingly, maximizing brand awareness is one of his fundamental objectives in this context. In the context of production control, minimizing order processing time may be desirable. When making plans for your vacation, you might pay attention to the amount of sun and with traffic planning you could try to keep cars out of the city center. However, one should be aware of the instrumental nature of these objectives; only unavoidable constraints of time, information, and resources can justify the restriction of an existing decision problem in which the underlying fundamental objectives will not be considered explicitly. In a given context, an objective which is essentially a means objective can be used as a "fundamental objective" if you have reason to assume that pursuing this objective will definitely be in line with your true wishes and preferences.

Even though aspects such as social responsibility and reputation have lately moved more into the focus of attention again, monetary objectives still prevail in the context of economic decisions. "Profit" is the most common fundamental objective in many decisions. On closer examination, profit is not an ultimate objective for anyone; rather, profit is a means to better achieve a number of other objectives including consumption (for shareholders), safety (for shareholders, employees and creditors) or self-affirmation (for managers). Nonetheless, profit can often be used as an appropriate fundamental objective since there is no doubt about its instrumental relevance to the true fundamental objectives.

Figure 3-1 illustrates how objectives that are fundamental in a narrow decision context can be means objectives in a broader context. For the design of a particular product, usability, design, manufacturing costs and recyclability are among the fundamental objectives. At least some of these would become instrumental, however, if the decision was made with respect to the whole range of products, thereby expanding the decision context. In this context, profit, market share, growth and liquidity could serve as fundamental objectives.

From the owners' perspective, however, these fundamental objectives are instrumental when deciding on whether to hold, sell or accumulate shares in their company. In the context of such a comprehensive decision, it is possibly a question of current consumption and future market value of the shares, as well as of self-affirmation as a successful entrepreneur and social prestige.

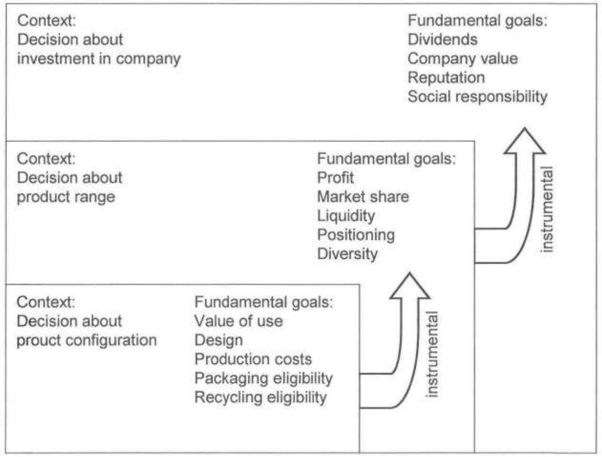


Figure 3-1: Fundamental objectives in a narrow context as means objectives in a broader context

As mentioned in Chapter 2, before making a decision you should check whether you are addressing the right problem. For example, is it appropriate to deal only with the design of a particular product or does it make more sense to combine this question with decisions on the full range of products? This would be an extension of the decision context that would almost certainly increase the set of possible alternatives.

3.4 Requirements for a system of objectives

We will call the entirety of (fundamental) objectives in a given decision context a *system of objectives*. With reference to Keeney (1992, pp. 82-86) we describe the most important requirements for a system of objectives below.

1. *Completeness.* The decision maker has to make sure that she takes into account all fundamental aspects of the consequences that she considers to be relevant. For example, it may be the case that socially unacceptable objectives remain unrevealed to other persons or sometimes even to oneself.

2. *No redundancies.* One should avoid having two or more objectives that actually mean the same thing or have an overlapping meaning. Otherwise, one runs into the danger of unconsciously attributing more weight to a certain objective than intended. When looking for a new flat, for instance, the objectives "good residential area", "quietness" and "no through traffic" would overlap, at least partly.
3. *Measurability.* The achievement of objectives should be measurable as accurately and unambiguously as possible. *As accurately as possible* means that the measurement should pertain to what is really important to the decision maker. *Unambiguously* means that the fuzziness of the measurement should be kept to a minimum. In Section 3.6.2, which deals with the choice/selection of appropriate attributes, measurability will be revisited and examined in more detail. For now, it suffices to note that the existence of appropriate attributes must not be ignored when determining the system of objectives.
4. *Preferential independence.* If possible, the decision maker should be able to express her preferences with respect to the attribute levels of a subset of objectives, independently of the attribute levels of the remaining objectives. The meaning of preferential independence will be explained in more detail in Chapter 6. For now, to obtain an intuitive grasp of the concept, imagine the following situation: a professional magazine is looking for an editor. Obviously, specialized knowledge and writing skills are two of the most important requirements for potential candidates. However, there is no preferential independence between these two objectives. They are complementary instead – each of the attributes becomes more important, the more the other attribute is fulfilled. The better the specialized knowledge, the more valuable the talent for writing. Both specialized knowledge without a talent for writing, and a talent for writing without specialized knowledge are worthless combinations with respect to the advertised position.

   Let us examine another example: a company is looking for a new production location; among the objectives are: "freeway access as near as possible" and "railway access as near as possible." If a more detailed analysis reveals that railway transportation and transportation by truck are interchangeable, there is no preferential independence between these two objectives. The better the freeway access of the location, the less important the railway access. In this case, the relationship between these objectives is substitutive.

   To prevent a common mistake, it should be noted that preferential independence has nothing in common with *statistical* independence. Within a given set of alternatives, there may be no correlation between the objective variables, e.g. between specialized knowledge and the talent for writing of the applicants; the attributes are thus statistically independent. However, this cannot be transformed into a statement about the preferential independence, i.e., whether a decision maker is able to evaluate one attribute without reference to the characteristic of the other attribute. Statistical independence is an objective fact, preferential independence in turn is generally a subjective question.

Preferential independence is an important characteristic of a system of objectives, because it allows an additive multi-attribute value function to be used. In simple terms, this means that the overall evaluation of an alternative can be calculated by adding up points for the different levels of each attribute. This is the simplest representation of preferences across multiple objectives that can be obtained (Chapter 6) and – due to its simplicity – is highly desirable.

If independence is not given in a certain decision context, it can often be achieved by redefining objectives. The abovementioned objectives for the traffic accessibility example could be replaced by the (more fundamental) objective of "minimizing shipping time."

5. *Conciseness.* The fewer the number of objectives included in a system of objectives, the less problematic the subsequent process of structuring the preferences and evaluating the alternatives becomes. Simplification can be achieved by aggregating objectives that lend themselves easily to aggregation. For example, when choosing a new production location, the objective variables "property costs", "construction costs", "transportation costs", "tax burden", etc. can be converted into a single objective variable such as the payout annuity.

   A simplification of the system of objectives can also occur if the search for alternatives has been completed and if it turns out that the available alternatives differ only marginally, if at all, with respect to a certain attribute. When looking for a new job, for instance, the salary is obviously an important objective from an a priori perspective. However, if it turns out that all available jobs are paid in terms of the same wage group, this objective can be eliminated from the system of objectives (as long as we have preferential independence, as mentioned in point 4 above), since in that case it becomes irrelevant for finding the best job offer. In other words: the objective is important, but it is irrelevant with respect to the available set of options.

3.5 Hierarchies of objectives

3.5.1 Higher-level objectives and lower-level objectives

It makes sense to use a hierarchical structure when developing the fundamental objectives in a certain decision situation. Objectives (in terms of fundamental objectives) can be broken down into *lower-level objectives* or aggregated into *higher-level objectives*. First, in such a hierarchical structure, it is easier to validate the completeness and lack of redundancies in the system of objectives. Secondly, dividing an objective into sub-objectives provides a means of describing the original objective more concretely and to enhance its measurability.

What is a lower-level objective? Health could be one of your fundamental objectives in looking for an appropriate recreational sport. However, this objective is

rather abstract and is inappropriate for the evaluation of some alternatives. How healthy is playing soccer, skiing or bodybuilding? "Health" can sensibly be broken down into lower-level objectives; each of the lower-level objectives then stands for a certain aspect of the higher-level objective. Following the sub-system of the human body, the objective of staying healthy can be broken down into lower-level objectives along the lines of a "healthy cardiovascular system", "healthy spine", "healthy digestive tract" and so on.

When buying a new car, passenger comfort is a complex and quite indefinite objective. Breaking this objective down into lower-level objectives like interior noise, headroom, legroom, and so forth helps make it more concrete and the lower-level objectives can be measured more precisely than the top-level objective of "passenger comfort." Please be aware of the following difference that sometimes causes some confusion: the concept of higher-level objectives and lower-level objectives should not be confused with the idea of fundamental objectives and means objectives. While the question of "Why do I care about this?" concerns the distinction between means objectives and more fundamental objectives, the distinction between higher-level objectives and lower-level objectives is made clear by the question "What exactly makes it important to me?".

3.5.2 Top-down- and bottom-up-procedure

A hierarchy of objectives can either be developed by a *top-down* or *bottom-up* approach. Top-down refers to developing the hierarchy of objectives "from the top to the bottom" by breaking down objectives into lower-level objectives. Starting with the comprehensive top-level objective (that could be characterized as "maximize the quality of the decision" or the like), you first have to ask which aspects are important for the overall quality of the decision; each of these aspects will then be broken down further if necessary. This procedure is appropriate if the decision maker already has a clear idea of the structure of aspects that have been identified as important for the decision. For example, when filling the vacant position of a head of department, some objectives can presumably be easily broken down further into lower-level objectives (Figure 3-2).

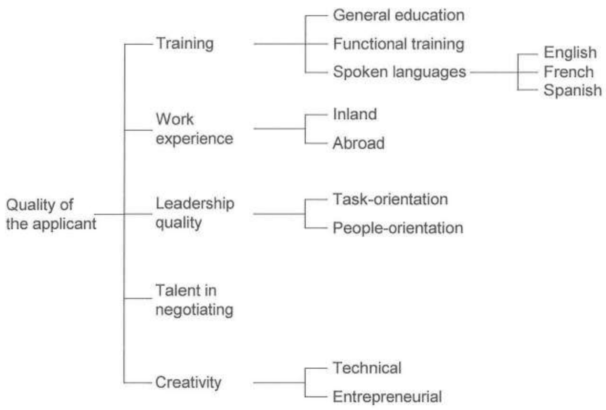


Figure 3-2: Hierarchy of objectives for filling a leadership position vacancy

When facing new types of problems, proceeding "from the bottom upwards" seems to be more appropriate. In this case, capturing all the relevant aspects as completely as possible is what is aimed for. For that purpose, it can be helpful to study the literature or relevant documents, consult experts and persons affected and so on. Having listed all possible objectives, one can try to aggregate objectives that belong together, to remove overlaps and to eliminate means objectives.

3.5.3 Development of a system of objectives

Usually, the *top-down* and *bottom-up* approaches are combined. The following example illustrates the procedure. A family decides to buy a dog and has to decide on the dog's breed. This kind of problem is new, as no family member has ever owned a dog before (or has ever collected any information on dogs). For a start, the family members set up the following list of subordinate objectives (*bottom-up* approach):

- should be fond of children,
- should guard the house,
- should guard the family during walks through the woods,
- should not be too big,
- should not be overbred,
- should be cute,
- should not be too expensive,
- should not need too much exercise,
- should not have long hair.

First of all, one should be clear on the question why each item in the list is an important objective. This makes it possible to identify means objectives and, if possible, to eliminate them. For some of the objectives, this is obvious: requiring the dog to be fond of children and cute, to guard the house and the family can easily be identified as fundamental objectives that are not subject to further discussion in the present context. (It is assumed that the problem is restricted to "buying a dog" in the present context.) But why should the dog not be "too big"? When discussing this issue, it turns out that behind this objective is the notion that bigger dogs need more exercise and thus are more time-consuming. The dog's size turns out to be a means objective that is inappropriate for several reasons: (1) the need for exercise is listed separately, (2) the relationship between size and need for exercise remains unclear (one can also think of big dogs that are quite inactive), and (3) the size has a positive impact on the protective function. Thus, "size" will be eliminated and the fundamental objective "minimize expenditure of time" has to be added.

The objective "not overbred" is a means objective, as well; what is actually meant by this objective is the fact that overbred dogs often tend to suffer from health problems; "health" is the fundamental objective.

Similarly, the demand for "short hair" is of an instrumental nature, since it is based on the conjecture that long hair means that more time will be needed for cleaning the carpet. This, too, is not a value, but a factual judgment (that might – in addition – be wrong) and will hence be eliminated.

After all of the above is taken into consideration, the family subsumes their objectives under five top-level objectives, as shown in Figure 3-3.

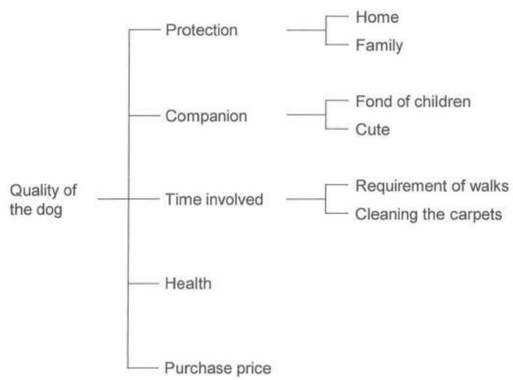


Figure 3-3: Original hierarchy of objectives for "buying a dog."

After this structuring of objectives, the objectives have to be checked in the reverse direction (i.e. *top down*) to check whether the breakdown into lower-level objectives is complete and non-redundant. It becomes clear to the parents that the dog should be not only a good companion for their children but also a good partner for themselves. They want to take the dog along when they go out jogging or on a bicycle tour, thus requiring the dog to have good stamina. The objective "low price" draws attention to the costs. However, dogs also differ in the costs of dog food and for veterinary treatment. These should be included as lower-level objectives under the objective "minimize costs". For reasons of simplification, however, it suffices to look at the aggregated "mean yearly costs" (including the purchase price divided by the life expectancy of the dog) when evaluating the alternatives at hand. This finally results in the hierarchy of objectives shown in Figure 3-4.

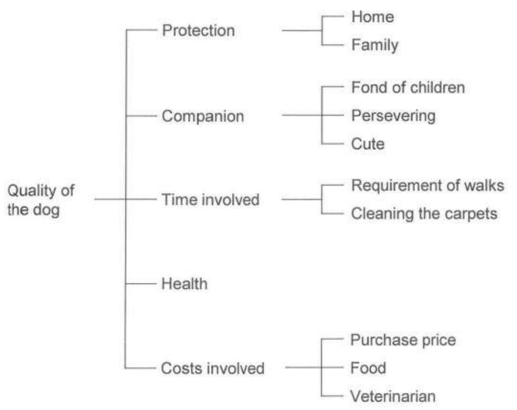


Figure 3-4: Final hierarchy of objectives for "buying a dog."

An objective hierarchy should not be confused with a *means-ends hierarchy*. The latter represents an assumed impact of means on the achievement of objectives, using intermediate steps if necessary. A means-ends hierarchy or – since a single means may affect more than only one objective – a "means-ends *network*" (Keeney 1992, pp. 87-92) includes factual knowledge about existing interdependencies. This can be the first step in developing a quantitative impact model. In addition, means-ends networks prove very helpful in generating sensible alternatives (we will come back to this point in the next chapter).

3.6 Attributes

3.6.1 Types of attributes

When determining the system of objectives for a given decision problem, appropriate attributes have to be identified to allow the measurement of the extent to which the objectives are achieved. Identifying appropriate attributes is particularly necessary for balancing the pros and cons of each alternative. In some cases "natural attributes" exist that can be used to measure the achievement of objectives, e.g. the attribute "profit in Euros" seems obvious if the objective is to "maximize profits". Analogously, it is quite obvious how to measure the power of a car engine, the print run of a newspaper or the domestic market share of portable TV sets. Often though, no natural attributes can be identified; for instance, it is difficult to quantify the aesthetic disfigurement of a landscape by overhead power supply lines (an effect which is intended to be kept as low as possible). The same applies to objectives like maximizing someone's quality of life or maximizing a company's reputation. In these cases, the decision maker either has to construct artificial attributes or rely on so-called proxy attributes that indirectly measure the degree to which the objective is achieved. The choice of attributes represents an essential part of the decision problem. Which alternative eventually turns out to be optimal is highly dependent on the attributes that have been chosen (Keeney 2007).

Natural attributes
As explained in Section 3.4, a good system of objectives should be measurable as precisely and unambiguously as possible. If a natural attribute is available, i.e. an objective variable naturally arises from its corresponding objective, these requirements can easily be met. It might thus even make sense to restructure a system of objectives or to further break down higher-level objectives into lower-level objectives to avoid artificial or proxy attributes. Unfortunately, however, this is impossible in some cases.

Artificial attributes
Artificial attributes (also often called constructed attributes) are usually constructed from a combination of several objective variables. In a study on evaluating different locations for a new water pumping station, the aesthetic and ecological disturbance of the landscape caused by power poles was measured by an artificial attribute "miles equivalent" (Keeney 1979). The actual length was multiplied by factors indicating the strength of disturbance. For example, the different factors meant

- Factor 1: Rural route traversing unpopulated rangeland; not visible from primary highways (3+ miles); does not affect any endangered species or unusual habitats; does not intrude on a "pristine" area.
- Factor 2: Urban route traversing populated areas, route traversing Bureau of Land Management or Indian-owned lands; or aesthetic intrusion on primary highways (more than 1 mile parallel to and within 3 miles of highway).

– ...
– Factor 5: Route traversing pristine undisturbed areas; or areas judged to be extraordinarily scenic.

For each location, the achievement of the objective "minimize disturbance caused by power poles" was determined by calculating the sum of miles for each route section, multiplied by its individual factor of disturbance.

Artificial attributes can also be formed by other combinations of sub-attributes and then already represent a small valuation model. A good example would be the "wind-chill" used in the Unites States which is a measure of perceived coldness and which combines the temperature and wind speed by an arithmetic function. When choosing between different medical treatments, the remaining life expectancy and quality of life may both play an important role. Obviously, the two attributes are not preference independent. The higher the quality of life, the more valuable each additional year of lifetime appears to be. To ensure independence within the system of objectives, it is common practice to aggregate both objectives into a single one. In the literature (e.g., Ubel et al. 2000, Hazen 2004), this is termed QALY (quality adjusted life years) and is calculated by multiplying additional lifetime by a special index that measures the change in life quality.

There is often no a clear distinction between natural and artificial attributes. In particular, artificial attributes that happen to be used very regularly may take on the nature of natural attributes as time goes by. For instance, many decision makers would accept the rate of price increase as a natural attribute of the objective "inflation". However, on closer inspection, this figure turns out to be a very complicated and artificial construct.

Proxy attributes

A second way of dealing with a lack of natural attributes is the use of so-called proxy attributes. These are attributes that do not directly measure the achievement of objectives; instead, the level of some auxiliary parameter, which is assumed to be directly connected to the objective, is determined. For example, the objective of customer satisfaction could be measured indirectly by counting the number of filed customer complaints per month. Employing proxy attributes often allows the use of simple measurement scales with attribute levels that are easy to determine. This avoids rather arbitrarily constructed artificial attributes. There is, however, the substantial drawback that the achievement of the objective can only be measured indirectly when using a proxy attribute. The suitability of a proxy attribute therefore depends strongly on the question of whether the assumed connection does indeed exist. Proxy attributes can arise in two different variations:

1. The proxy attribute is an *indicator* of the achievement of an objective.
2. The proxy attribute can be characterized as a *means* of achieving the objective.

The aforementioned example of customer complaints as an indicator of customer satisfaction belongs to the first variant. Likewise, the existence of certain plants or animals could be considered an indicator of environmental pollution. In economically relevant decisions, companies often aim for certain balance sheet ratios that do not have much of a meaning per se but are often considered by banks as an indicator of the company's creditworthiness. When choosing between different candidates for a professorship, the number of publications in highly ranked journals plays an important role. The actual objective is to hire a professor who will be a productive researcher in the future; the track record from the past only serves as an indicator. The problem with all these indicator variables is that they are not perfectly correlated with the objective variable. Thus, levels of achievements of objectives might be assumed that do not really exist.

Examples of the second variant would be the emission of oil into the ocean by oil rigs or the size of the ozone hole as proxy attributes for the objectives "conservation of marine life" or "protection against harmful UV-radiation". The quality of a practical training may be measured by the number of lessons, the healthy impact of a diet by nutritional facts regarding fat, cholesterol, sugar, etc. A private equity company could measure the potential of a target company by looking at the proxy attribute "last year's expenditure on research and development". A customer looking for a new hairdryer could measure the "drying performance" by means of the proxy attribute wattage.

The particular problem with instrumental proxy attributes is apparent; it is unclear how strong the causal impact of the proxy attribute is on achieving the underlying objective. While one can reasonably assume a rather strong relation in some cases (like the influence of wattage on drying ability), the suitability of proxy attributes is rather dubious in other cases (as in the aforementioned case of measuring the potential of a target firm by its R&D expenditure). However, due to a lack of alternatives, the use of such problematic proxy attributes sometimes cannot be avoided.

3.6.2 Desirable properties of attributes

In the course of this chapter, some desirable attribute properties as well as some problematic attribute characteristics have been addressed. The desirable properties will now be reconsidered in order to provide an explicit catalogue of criteria for choosing attributes. Following Keeney and Gregory (2005), we identify five properties: Direct, Operational, Understandable, Comprehensive and Unambiguous.

Direct

An attribute should relate as directly as possible to the achievement of a given fundamental objective. If, for instance, a company aims for a reduction of costs of its bloated administration, the attribute "number of administrative employees" measures the achievement of this objective very indirectly. One could think of restructuring measures that effectively reduce the number of employees but at the same time, induce higher costs caused by some additional highly-paid jobs that had to be created. Therefore, directly measuring the actually relevant attribute "(personnel) costs of administration" would be the superior choice.

Operational
An attribute is useful in solving a decision problem only if the attribute's level can be determined with reasonable reliability and with a reasonable amount of effort for all existing alternatives. For example, it might be theoretically sound to measure health hazards resulting from various cell phone models' radiation by using an artificial attribute which takes into account the probabilities of certain diseases occurring after a certain time of use. However, not only would the effort required to determine the attribute's levels be unreasonably high, but the results would also be quite uncertain, being based on the current state of knowledge, rendering this attribute impractical.

Understandable
The decision maker should be able to fully understand the meaning of the attribute used and in particular, understand its various levels. This is often not the case, e.g. much of the data published by technical magazines in the context of Hi-Fi or personal computer reviews is not readily understandable and thus virtually useless to the layman. Understandable attributes are relevant, not only for the communication of advantages and disadvantages of the various alternatives but also important to the individual decision maker, allowing him to balance versatile combinations of the attributes' levels. As an example, would you be instantaneously able to say what deterioration in the distortion factor of a high-end-amplifier you would be willing to accept in exchange for a certain amount of cost savings?

Comprehensive
Attributes should be chosen such that they allow a consideration of all relevant consequences with respect to the achievement of objectives. If, for instance, various transport policy measures have to be evaluated with respect to the objective "safety of traffic participants," the attribute "number of fatal accidents per year" is not sufficiently comprehensive as it does not account at all for severe accidents without fatalities. A more suitable attribute would include the number of injured people as well as the death toll and combine them with the severity of an accident weighted by various criteria.

Unambiguous
Attributes should be chosen such that it becomes unmistakably clear from their levels to what degree the objective was achieved. Therefore, very broad categories for measurement should be avoided. School grades, empty phrases in job reference certificates, the number of stars in book reviews, or the number of "forks" in a restaurant guide are all relatively vague. By contrast, the maximum speed of a car (mph) or the starting salary of a position ($/year) can be measured quite precisely. Requiring unambiguous attributes should not mistakenly be interpreted as requiring the attributes' levels to always be certain. The salary you will earn next year, for instance, can be measured quite exactly, but can be uncertain from today's perspective, due to possible bonuses. Attention should be paid to avoid choosing an attribute which aggregates the uncertainty in a way (e.g. by just looking at expected values of the salary) but ignores important characteristics (e.g. the variance of the salary) of the consequences.

3.6.3 Determining suitable attributes

In many cases, none of the available attributes will meet all of the above mentioned criteria equally well. The search for an optimal attribute can thus be considered a small but separate decision problem, in which the pros and cons of the attributes in question have to be balanced with respect to the abovementioned criteria.

If a natural attribute exists, this will usually be the best choice, particularly if it is not only direct but also operational and comprehensive. However, even with natural attributes, some scope/uncertainty of definition remains – a fact of which the decision maker (or the decision analyst) should be aware. For instance, Keeney (2007) shows that, in practice, it can indeed make a difference whether costs are measured and stated in terms of millions or in terms of thousands of dollars. This can cause a different perception of the size of differences. Indeed, the difference between costs of €7.1bn and €7.9bn seems rather insignificant in the context of bailout plans in the recent financial crisis which changes dramatically if that difference is expressed as €800,000,000.

Another interesting example arises when you think about choosing an appropriate attribute for measuring the mileage of cars. The ratio of liters per 100 km (l/100km), for instance, is so common in Germany that it is actually considered a natural attribute for measuring the mileage. A car *b* with a mileage of 7 l/100km is likely perceived to be midway between car *a* with a mileage of 5 l/100km and a car *c* with a mileage of 9 l/100km. Using the attribute miles per gallon (mpg) instead, which is prevalent and considered a "natural" attribute in the US, things look much less "symmetrical"; *b*'s consumption of 33.6 mpg would be just slightly superior to *c*'s consumption of 26.1 mpg but considerably inferior to *a*'s consumption of 47.0 mpg.

Nevertheless, such problems related to the use of natural attributes should be considered rather minor and can be eliminated by a sound and systematic derivation of a value function, as described in Chapter 5.

If no natural attributes can be found to measure the achievement of objectives, one should first try to construct an artificial attribute that is at least acceptable with respect to the aforementioned criteria. Unfortunately, artificial attributes often turn out to be difficult to understand so that considerable attention should be paid to this issue when constructing such attributes. Hence, artificial attributes that result from weighted averages of various sub-criteria are often better understood than attributes that result from complicated combinations of key figures. Only if it is not possible to generate an artificial attribute that would be acceptable – or if there is not enough time to think thoroughly about the construction of such an attribute – should the decision maker resort to using proxy attributes and thereby accept a very indirect measurement. If proxy attributes are about to be used, you should always ask yourself whether the instrumental relation is sufficiently pronounced to keep it out of the modeling of objectives and rather incorporate it into the impact model instead. Suppose, for example, that it is possible to predict with reasonable accuracy the effects of the proxy attribute "carbon dioxide emissions" on changes in the global temperature and also to predict the effects on the more basic objective; namely the influence of temperature changes on human beings, animals and plants. In this case, alternatives (such as statutory limitations or taxation of emissions) could be measured directly by the consequences that are of real interest. Only if the consequences are too vague to be modeled, but if it is nonetheless expected that holding constant or increasing the current level of carbon dioxide emissions could have catastrophic consequences, would it become necessary to use these emissions as a proxy attribute. Note that it is somewhat paradoxical that, on the one hand, a strong impact of the proxy attribute on achieving the fundamental objective is assumed, while, on the other hand, this impact is too vague to be captured in an impact model (even a probabilistic one).

Questions and exercises

3.1

One of your friends wants to buy a used car. Although he does not have a clue about decision theory, he has prepared a list of criteria:

- low price
- minimum of rust
- low kilometer reading
- preferably new
- good condition
- catalytic converter
- tidy owner
- low fuel consumption
- large loading volume
- side collision protection.

Do you think this is a useful aid or is there a way to improve this system of objectives? If so, how?

3.2

Imagine you are undertaking a one-year expedition to the Antarctic. You will have a lot of time available that you would like to use to learn another foreign language – you are already proficient in German and French. CDs and textbooks are available for many languages. Your problem is choosing the language. Create a system of objectives for this decision. Organize it hierarchically if possible. Test whether it fulfills the requirements given in the text.

3.3

You study at an overcrowded university and have been voted student representative of your faculty. A committee wants to explore options for changing the conditions themselves. The objectives agreed upon are: (1) improvement of the studying conditions for students; (2) a teaching load reduction for professors and assistants; and (3) at least no decrease in the quality of education.

(a) Concretize the objectives by forming measurable lower-level objectives. Try to give every lower-level objective an attribute by which the level of achievement can be measured.

(b) Check whether or not the attributes that you determined in (a) meets the criteria that have been discussed in this chapter (direct, operational, understandable, comprehensive, unambiguous).

3.4

Noise pollution impacting on residents in the immediate vicinity is an important factor in the site selection for a new airport. Construct an attribute that includes all aspects of noise pollution.

3.5

An army wants to introduce a new type of rifle. The alternatives differ in reach and accuracy. Do you think these two target variables are mutually preference independent? If not, how can the preference dependence be corrected?

3.6

The personnel manager appraises graduates, among other things, by their final grades, but also by their study durations. In order to compare applicants with each other, he has developed the following points chart:

	Mark 1	Mark 2	Mark 3	Mark 4
10 terms	100	80	65	55
12 terms	80	66	53	42
14 terms	65	55	45	35

Characterize the personnel manager's preference for both attributes.

3.7

In a comparison test of different vehicles, gas consumption is calculated as the average consumption in city traffic, on highways and on expressways. The new Lopu achieves 9 l/100km, 7 l/100km and 5 l/100km. The Pundu performs at 8.8 l/100km, 6.8 l/100km and 4.8 l/100km.

(a) With regards to the attribute "low gas consumption", which one is the better car of the two?

(b) Assume this comparison would not have been carried out in Germany, but in the US instead. Thus, gas consumption would have been calculated in miles per gallon. Describe why the observed effect occurs.

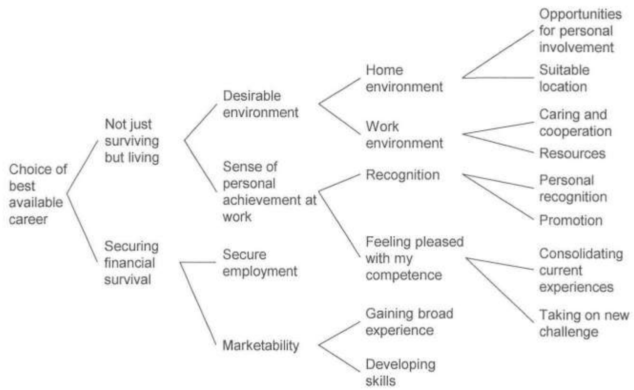
Case Study: Career Planning Support at ICI

Source: Wooler (1982), pp. 343-351.

The British consultancy ICI developed a decision analytic system to support ICI employees with their career decisions. The system was intended to function in such a way that interested employees ("clients") would be able to use it without any external aid, and was intended to serve to facilitate the structuring of the clients' objectives and further aid evaluation of different career options.

One module of the system was HISTRA ("Hierarchical Structuring Aid"), a program developed by ICI. The intention of HISTRA was for each client to become clear about her personal objectives relevant for career planning and to express their subjective importance.

For this purpose, individual objective hierarchies were developed in 17 interviews with individual employees. On the basis of these interviews, the following common objective hierarchy was constructed and implemented in HISTRA.



In a HISTRA session, the program would explain to the client the objective hierarchy and ask her to evaluate the importance of single objectives using a simple relative weighting mechanism. The information gathered concerning the client's preferences is then the input for a further program named MAUD (Humphreys and McFadden 1980) that helps the client evaluate her career options.

Chapter 4: Generating and preselecting alternatives

4.0 Summary

1. The application of cause-effect analysis naturally leads to alternatives. Formal impact models and means-ends networks may generate a multitude of useful alternatives.
2. Visualizing an ideal alternative may stimulate creativity on how to approach such an alternative.
3. Expanding the decision context by taking into account more fundamental objectives may yield innovative, previously unrecognized options.
4. In many cases, the construction of a good solution can be broken down into various modular partial solutions that may simplify the decision problem.
5. By developing multi-level alternatives that include a number of possible reactions to future events, one can obtain more - and often better - alternatives than by generating only single-level alternatives.
6. Creativity techniques like brainstorming and the Nominal Group Technique can enhance or support the development of alternatives in groups.
7. Identifying and eliminating bad alternatives is important if the number of options becomes very large. This may mean that the decision maker has to preselect alternatives without having completely evaluated them beforehand. Restrictions and aspiration levels are popular but problematic methods. Under certain conditions, it is possible to preselect alternatives according to the dominance criterion.

4.1 The generation of promising alternatives

The purpose of the decision process is to find the best alternative that can be achieved. In some cases, the complete set of alternatives is "given", e.g. in a polling booth. In many cases, however, the search for and generation of alternatives is difficult and complex.

Which alternative is the best depends on the decision maker's fundamental objectives. Searching for and generating new alternatives should therefore be a goal-oriented process, with the objectives indicating the direction in which to look for better alternatives. Even if the system of objectives is still preliminary and could be modified until the final decision (in particular due to the emergence of new alternatives), the objectives are the most important guideline in generating additional possible actions. To put it differently, the search for options is not very promising if you do not know what you want. The system of objectives should thus be as clear as possible before time and money is spent on generating options.

In principle, the more time you have available the better the options you can generate become; strong time pressure is a very unfavorable condition. Examples

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_4, © Springer-Verlag Berlin Heidelberg 2010

of unfortunate reactions to problems that could have been foreseen were given by the Swiss pharmaceutical company Sandoz after the contamination of the river Rhine by toxic chemicals in 1986 or by Exxon after the supertanker accident in Alaska 1989, which caused oil contamination along the coastline for hundreds of miles. It is helpful to prepare early for decisions that are foreseeable.

A systematic search for alternatives is often omitted in practice. All too often, only a few proposals are generated and focused on in the discussion that follows. The individuals involved tie themselves down to some alternatives and do not like it if a new alternative appears that needs to be considered as well.

The process of reviewing and, if necessary, eliminating alternatives is closely related to the generation of new alternatives.. It does not make sense to simply increase the number of options. The final goal of the process is, in most cases, to end up with only a single - the best - alternative. Hence, increasing the number of alternatives is just an interim stage before reducing them again. Finding or generating new alternatives is promising only if there is a good chance of finding an alternative that turns out to be better than the best of the already existing options. However, this requires an evaluation of the quality of alternatives – at least roughly – before making a final selection. We face the problem of making a preselection: deciding whether a specific alternative is promising enough to justify a more detailed analysis.

In the remainder of this chapter, we first discuss some systematic approaches to generating alternatives. What follows is a brief description of creativity techniques that were developed for gathering ideas in groups. Finally, we discuss methods of preselecting alternatives.

4.2 Cause-effect analyses

4.2.1 Creating alternatives by means of a quantitative impact model

Hypotheses on cause-effect relationships can often be a source of ideas concerning possible alternatives. A physician wants to reduce a patient's blood pressure; he has certain ideas about the means by which this can be achieved. Economic policy makers have hypotheses on the effect of raising the central bank discount rate on inflation and the balance of payments.. Sales managers are able to estimate the marketing budget that is necessary to raise the market share by one percent. This (albeit uncertain) knowledge about causal relationships can be used to derive potential alternatives.

From previous chapters, you should be familiar with impact models, event trees, cause trees, influence diagrams and means-ends networks; all these concepts can be used to support causal analyses. In this subsection, we present an example of the use of a quantitative impact model, and in the next subsection, an example of the application of a means-ends network.

Impact models, as defined in Section 2.4, deterministically model the combined effect of both alternatives and environmental influences in the attribute space. If there is only a single objective variable, or if the multiattribute preference model already exists, it might be possible to identify the optimal alternative by using the impact model. If the decision maker pursues multiple objectives and has not yet determined his preference model, the impact model can at least support him in finding alternatives that are better than others.

Consider the following example: a manufacturer of a branded product wants to find the optimal marketing mix for a certain product; we restrict the mix to the product price and the advertising budget. Suppose the company's system of objectives consists of the monthly profit, monthly quantity of sales, value-based market share and price continuity. Sales and market share have been included in the system of objectives because they are not only considered a means of generating profit but also play an independent role in the further development of businesses.

Suppose the marketing planners have formed some expectations of customer reactions with respect to the price p and the advertising budget B that models the sales quantity x as a function of price and the advertising budget as a sales function of the form $x = x(p, B)$. In addition, there is a defining equation for the profit, including turnover and costs: $G = px - K$. Define the attribute "market share" by $px / (px+U_S)$, with U_S representing the aggregated turnover of all other suppliers. Price continuity is measured by the absolute difference between the current price (€15) and the new price, $| p - 15 |$.

These functions represent the impact model that transforms the decision variables p and B into the objective variable. Such a model allows us to simulate various combinations of p and B and thus identify viable alternatives. Combining monthly advertising budgets from 300 to 900 (thousand €) with prices of €18, €16 and €14, for instance, results in values for the objective variable as shown in Table 4-1.

Table 4-1: Different marketing mixes based on a quantitative impact model

Decision Variables			Objective Variables		
Advertising budget (in €1,000)	Price (in €)	Sales (in 1,000 units)	Profit (in €1,000)	Market share (in %)	Price difference (in €)
300	18	147	879	11.7	3
300	16	190	837	13.2	1
300	14	232	627	14.0	1
400	18	198	1.182	15.1	3
400	16	254	1.126	16.9	1
400	14	311	843	17.9	1
500	18	243	1.443	17.9	3
500	16	312	1.374	20.0	1
500	14	382	1.027	21.1	1
600	18	282	1.654	20.2	3
600	16	362	1.573	22.5	1
600	14	443	1.171	23.7	1
700	18	313	1.803	22.0	3
700	16	402	1.714	24.3	1
700	14	492	1.267	25.6	1
800	18	335	1.880	23.2	3
800	16	431	1.784	25.6	1
800	14	526	1.306	26.9	1
900	18	346	1.870	23.8	3
900	16	445	1.771	26.3	1
900	14	544	1.276	27.6	1

A closer inspection reveals that advertising budgets between 300 and 700 can be ruled out. Each consequence linked to one of these budgets is worse, or at least not better, than one of the other consequences with respect to all objective variables. For instance, $B = 600$ and $p = 18$ is inferior to the alternative $B = 700$ and $p = 18$, because its levels are worse for three of the attributes (sales, profit, market share), , and its level is the same for the remaining attribute (price difference),. Only the remaining six alternatives (in the rectangle) are worth considering further. Each of these alternatives yields advantages and disadvantages compared to each of the other alternatives. The other options (not included in the rectangle) are all *dominated*, in the sense that each of them is outperformed by another alternative (from the set of dominating alternatives) in all respects (or is outperformed in at least one dimension and is not superior in all other attributes). Only *non-dominated* alternatives are shortlisted. Which of them will finally emerge as the best one depends on the multiattribute value function which remains to be specified (see Chapter 6).

By analyzing the cause-effect hypotheses implied by the impact model, we were not only able to generate a number of viable alternatives but also identified several alternatives that were not viable.

4.2.2 Alternatives as a combination of measures

The conjectured instrumental relationships – illustrated in a means-ends network – are a useful starting point for discovering all the promising combinations of measures.

Assume, for example, that absenteeism has become a big issue in the production department of a company. More and more frequently, a significant part of the production capacity cannot be exploited because of the absence of staff. Confirmed delivery times cannot be met anymore, customers complain and orders are cancelled. The task force that was installed to compile suggestions on how to approach the problem contemplates possibilities for reducing the absenteeism rate (it is assumed that the actual frequency of illness is exogenous and cannot be changed). The objective is thus to minimize absenteeism caused by employees simply skipping work; another objective variable is the cost of the measures to be taken.

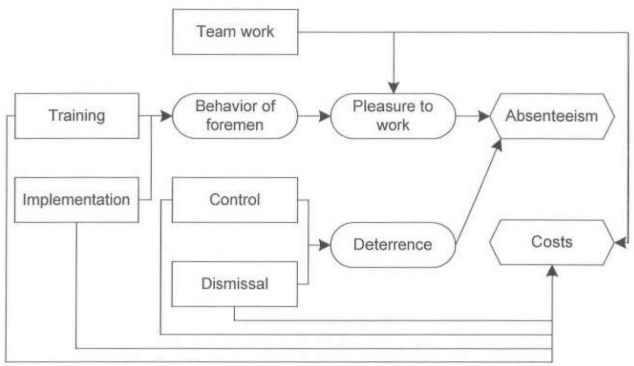


Figure 4-1: Means-ends network for the objectives "minimize absenteeism rate" and "minimize the cost of associated measures"

Thinking about ways and means to achieve these objectives results in a means-ends network, as illustrated in Figure 4-1. The rectangular boxes stand for *measures* that are intended to have an effect on the objectives (in hexagons). The

rounded boxes contain the means objectives, which have an indirect (the behavior of foremen) or direct (the joy of working, deterrence) effect on the fundamental objectives, as long as the underlying effect hypotheses prove to be correct. The arrows in the figure indicate the direction of the effect; they must be specified in more detail by an impact model. If the effects are uncertain, the uncertainty has to be modeled as well. However, this is beyond the scope of this chapter.

The absenteeism rate is thought to be influenced partly by low joy of work. The hypothesis is: the greater the joy of work, the less the employees tend to stay away. A possible measure for enhancing the joy of work could be to replace assembly-line work with teamwork since the latter offers more opportunities for communication and creates more work variation and decision-making situations. It is also less exhausting in general. In addition, the working atmosphere often suffers from the behavior of some foremen who are perceived as dictatorial and unfair. Possible means of improvement could be to conduct special training for the foremen or to redeploy them to some other position where their poor leadership abilities are less harmful.

Alternatively, one could try to achieve the objective through deterrence. Absence from work could be recorded for each employee individually ("control") and may thus lead to fears of dismissal; one could even fire one or two notorious "sinners" as a warning.

What actions can be derived from the means-ends network? We can try to construct all possible combinations of measures. Each combination of measures then represents a potential alternative. If we assume that n measures could be realized in any arbitrary combination, this results in 2^n possible alternatives, including the null alternative of doing nothing. In the present example, this amounts to $2^5 = 32$. In the case of $n = 10$, the number of alternatives would increase to 1,024, too many to analyze them all.

In most real decision situations, not all possible combinations will make sense. For example, after introducing teamwork, the role of the foremen has to be redefined, making obsolete measures aimed at improving their leadership ability . This procedure eliminates twelve of the alternatives in which "teamwork" was combined with "training" or "redeployment". Combining "training" and "redeployment" does not make sense either, so that another four alternatives can be removed. The remaining sixteen alternatives are listed in Table 4-2.

Table 4-2: Some sensible-looking alternatives for reducing absenteeism

1	Training	9	Redeployment + Dismissal
2	Redeployment	10	Control + Teamwork
3	Control	11	Control + Dismissal
4	Dismissal	12	Dismissal + Teamwork
5	Teamwork	13	Training + Control + Dismissal
6	Training + Control	14	Redeployment + Control + Dismissal
7	Training + Dismissal	15	Control + Dismissal + Teamwork
8	Redeployment + Control	16	Doing nothing

We assume that, by using a less systematic approach - as is common in practice - only a fraction of these alternatives would have been considered further.

4.3 Ideal alternatives

The tendency to look for new solutions in close proximity to existing alternatives may hinder creativity. Companies usually try to remedy problems by slightly modifying prevailing procedures (Cyert and March 1963, pp. 121-122). This strategy, however, does not apply only to corporations. Professors who are asked to draft a new version of the examination rules all too often fall back on the rules that already exist and try to improve some aspects of them. (At most, they also consult two or three examination rules from other universities for comparison.) The tendency of minimizing effort by merely making some "adjustments" instead of contemplating fundamentally new opportunities is understandable but not always the right way to proceed.

Thinking about alternatives under multiple objectives can be simulated by imagining an *ideal* solution. Based on a well-defined system of objectives, one would conceive what an alternative should look like, ideal in the sense that the decision maker is completely satisfied with the result. This would be an alternative that shows the optimal conceivable level with respect to all objective variables. Ideal solutions can seldom be reached, but they can be a useful starting point for discovering new alternatives.

At a restaurant, instead of studying the menu helplessly because you cannot find a dish that appeals to you, you could consider what kind of dish you would like best right now, i.e. you generate an alternative that would be perfect for you in that situation. Afterwards, you can try and talk to the chef de cuisine. Perhaps he can recommend something that comes close to your ideal; supposedly, you will be happier with the result than what you would have obtained when restricting yourself to what is on the menu.

Often, existing alternatives come off pretty well with respect to one or more objectives but are really poor with respect to others. One would come very close

to the perfect solution if it was possible to combine the best attribute values of various alternatives.

For example, assume you are looking for a new secretary. On the one hand, you expect her to be sociable and to get on well with people, because there will be customer contact in the morning. On the other hand, you expect her to know how to use a computer, including various complex applications. Unfortunately, there is no candidate among the applicants who meets both requirements equally well; however, there are two who meet at least one of the required qualifications. Hiring both of them for a part-time job could be a creative solution, thereby combining the best qualities of both.

If an alternative has been identified that turns out to be good in several respects, but very bad with respect to other attributes, it may be worthwhile thinking about how to improve this alternative with regard to its weaknesses. When looking for a new location for a storage power station in New Mexico, ten alternatives had been taken into consideration (Keeney 1979). One of them featured the lowest cost and a minimum disturbance of the landscape through power cables, but was disadvantageous with respect to environmental damage to the riverside biotopes. It was an almost perfect solution; thus, it was modified such that another nearby valley was chosen for the lower basin (Keeney 1992, p. 212). As a result, riverside damage could be avoided and an almost perfect solution was found.

4.4 Context enlargement

As mentioned above, when developing the objectives for a specific decision, it is always worth wondering whether the context should be extended by switching from some initially created objectives towards a set of more fundamental objectives. Such an extension often brings to light some new, previously unconsidered alternatives.

We have already presented a number of examples. One of them concerned a company whose products caused customer complaints at a rate that management perceived as unacceptably high. Initially, production quality improvements like intensified quality controls or quality bonuses had been proposed. However, by replacing the (means) objective "errorless production" with the more fundamental objective of "raising profits", some other alternatives were brought into focus, such as the external procurement of critical parts instead of in-house production or a price reduction for the product.

4.5 Decomposition of tasks

For some decision situations one has to *find* the alternatives (like a CEO, an apartment or a used car); in other situations, though, one has to *invent* them (a text, a certain behavior or a product design). The invention of alternatives is often a constructive task that can be hierarchically structured. In organization theory, this concept is known as "task analysis" (see e.g. Frese 1987). One aim of task analysis is to obtain a complete overview of all the elementary tasks that can be sensibly combined afterwards and assigned to different people (task synthesis).

Generating one or more complex alternatives through the decomposition of tasks is the opposite of the procedure described in Subsection 4.2.2 (combination of measures). The task of finding as good an alternative as possible is broken down into the design of individual modules. Within each module, one looks for one or a small number of good solutions. Afterwards, the module solutions are combined to obtain one or a few alternatives. Although there are surely interdependencies between the different modules, some decoupling is usually possible. During the design process, it may turn out that some modules meet the objective requirements to a greater extent than others. One can then consider the "good" modules as finished for the moment and concentrate on improving those that are not yet satisfactory. The alternatives that come up for the final evaluation might then only differ in the sense that there may be multiple variants for some of the modules. The evaluation of a potentially huge number of alternatives can thus be greatly simplified by a series of (pre-)decisions about some of the modules.

Suppose you want to prepare dinner for some friends. Your repertoire includes ten appetizers, twelve main courses and six desserts. Instead of comparing all of the 720 possible combinations, you could rather determine the one or two best appetizers, the one or two best main courses and the one or two best desserts first. This results in one to eight complete menus.

By decomposing tasks, it also becomes possible to distribute the complete task among several people (e.g. experts) and to work in parallel on the different subtasks.

As another example, think about the task of finding the best possible wording when designing a general partnership agreement. At first, we consider all the necessary parts of the contract, like

- Firm and registered office,
- Kind of business,
- Partners and cash contributions,
- Making of decisions by the partners,
- Procuration,
- Profit assessment and distribution of earnings,
- Withdrawal of capital, etc.

Each of these parts can be broken down further, e.g. with respect to the "making of decisions", the following aspects should be specified:

- Mode of decision making (in written form or at the partners' meeting),
- Required majorities,
- (Presence of a) quorum,
- Agent's authority, etc.

One can further subdivide the required majorities with respect to regular issues, modifications of a contract, acceptance of new partners, liquidation of the business, etc.

Assume that the partners divide up the task of drafting the contract. Each of them takes on some modules. Afterwards, they forward the various drafts to each other. After some discussion and revisions, the main part of the memorandum of partnership is "fixed". Some aspects are still subject to discussion, so some counter-proposals will also be outlined. The final evaluation and decision are confined to these alternatives.

4.6 Multi-level alternatives

Multi-level alternatives have been introduced in Subsections 2.2.3 and 2.7.4; a multi-level alternative is a sequence of conditional instructions. For decisions under uncertainty, the number of promising alternatives can often be increased by expanding the decision to multiple levels. Of course, the procedures described in Subsections 4.2 to 4.5 and the generation of multi-level alternatives are not mutually exclusive.

In many cases, multi-level alternatives are necessary, because single-level alternatives would be senseless from the start. Introducing a new product, for instance, is exposed to some risk; if the product turns out to be a flop, it would be foolish to risk complete financial ruin by adhering to it. Possible reactions to a flop, i.e. alternatives like product improvement, increasing the marketing budget or shutting down production should be taken into account from the start. Just as it seems a good idea to think one move ahead of your opponent in a parlor game, including a second or third decision step will generally not worsen the set of alternatives but often improve it.

Look at the following example. A food producer, Nibblingfactory Inc., is forced to dispose of 20 tons of ready-made meals after an incident of salmonella contamination. The company suspects a particular primary product that is purchased from a rural producers' cooperative to be the cause. The management board of Nibblingfactory Inc. considers suing the supplier for damages of €100,000.

The evidence, however, is somewhat inconclusive so the outcome of the case remains uncertain. If it loses the case, the company has to pay the legal costs (approximately €10,000). Thus, there is some incentive to agree on an out-of-court settlement. The board takes into consideration a €75,000 settlement offer and expects the supplier to react with one of the following three alternatives:

- acceptance of the €75,000 settlement offer,
- refusal of any compensation for damages,
- a counter offer, presumably around €50,000.

The situation can be illustrated by the decision tree shown in Figure 4-2.

Figure 4-2: Decision tree of Nibblingfactory Inc., including both one single-level and four two-level alternatives

However, in case that the supplier refuses the settlement offer, it seems quite natural to include the other options for further analysis. If the supplier refuses any payment, Nibblefactory Inc. could still take legal action. Another option would be to submit a reduced settlement offer of, say, €50,000. The board is firmly convinced that the producers' cooperative would agree to this settlement, if only to keep an important customer.

In the case where the supplier responds to the first settlement offer with a counter offer of €50,000, Nibblingfactory Inc. is also in a position to decide on either filing a suit or accepting the counter offer.

These considerations result in the situation illustrated in Figure 4-3. By now, the following five strategies are available to Nibblingfactory Inc.:

a File a claim for 100

b Offer a settlement of 75;
If supplier refuses: sue for 100
If supplier offers 50: sue for 100

c Offer a settlement of 75;
If supplier refuses: sue for 100
If supplier offers 50: accept

d Offer a settlement of 75;
If supplier refuses: offer a settlement of 50
If supplier offers 50: sue for 100

e Offer a settlement of 75;
If supplier refuses: offer a settlement of 50
If supplier offers 50: accept.

Among others, the evaluation of these strategies depends on the probabilities associated with the uncertain events, i.e. on the prospect of success in court and the probabilities associated with the supplier's reactions to the settlement offer of €75,000.

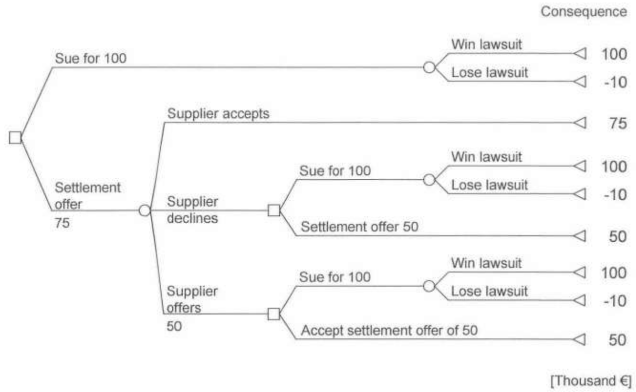


Figure 4-3: Decision tree of Nibblingfactory Inc., including both one single-level and four two-level alternatives

4.7 Creativity techniques for groups

The methods of generating alternatives discussed so far are based on an analytical-systematic approach. In addition, a number of procedures have been developed that aim to support a decision maker's intuition. They try to induce spontaneous ideas and overcome thinking barriers. Usually, creativity is hindered by a decision maker's tendency to look for a problem solution very close to the status quo, minimizing cognitive efforts and the risk of failure. In addition, group-psychology effects, such as fear of criticism (in particular from higher-ranked group members) or the tendency to conform in a coherent group, hinder the development of new ideas.

At this point, we will mention just two of the most common methods suggested for group discussions (see Schlicksupp 1999 for a comprehensive presentation and Hauschildt 1994 for a critical review).

4.7.1 Brainstorming

This pervasive method (Osborn 1963) is considered to be quite successful. The group discussion should abide by the following simple rules:

1. Each participant spontaneously expresses the ideas that come to his mind, regardless of how abstruse they may seem. All of them are noted.
2. Any form of criticism or insulting remark is prohibited.
3. The participants are also encouraged to link their remarks to the ideas of the previous speakers and to refine them.

4. As many ideas as possible should be generated.

Brainstorming is used to create an atmosphere in which everyone dares to express some ideas, even if they are rather crude. The spontaneity of others is intended to be inspiring. The more unusual the proposals that have already been put forward, the more uninhibited the new ideas that will be expressed. Typically, the participants loosen up during a brainstorming session; at the beginning, rather conventional suggestions will be made, but over the course of the session, more and more unusual and innovative ideas emerge.

4.7.2 Nominal group technique

The aim of NGT (Delbecq et al. 1975) is to avoid some participants being almost excluded from the discussion while some dominant group members take control. During a normal and unstructured discussion, everybody is busy speaking or listening; in general, there is not much time for thinking. The process of generating ideas with NGT is as follows:

1. A period of silence, in which each participant develops his or her own ideas and writes them down.
2. One after the other, each participant presents one of his ideas. The ideas are recorded on a blackboard. This phase is completed after all the individual lists of ideas have been noted.
3. Discussion and clarification of all the ideas that were listed on the blackboard.

This procedure gives all participants the opportunity to reflect on their own ideas, unaffected by others and without being disturbed or subjected to stress. A focus on already existing proposals and a tendency towards premature consensus can thus be avoided. All ideas are made available to the entire group and have an equal chance of success. The discussion in the third phase is less personalized than in unstructured processes.

4.8 Preselection of alternatives

4.8.1 The need for preselection

The generation of alternatives is not an end in itself. Each alternative that has to be analyzed in more detail causes some effort. When making a decision under certainty, attribute levels for all objective variables and alternatives have to be determined. This may require extensive information retrieval, e.g. collecting a large body of technical and economically relevant data about the existing options. If you are looking for a used car, you won't look at hundreds of offers in great detail, nor would you want to conduct dozens of job interviews when looking for a new secretary.

In the context of uncertainty, the decision is even more complicated as a single alternative may lead to different consequences. In addition, the uncertainty has to be quantified by assigning subjective probabilities to the consequences.

To put it differently: the more alternatives you have, the greater your need for pre-selection becomes. For it to work effectively, the pre-selection has to involve less effort than the ultimate evaluation of the remaining alternatives. This is a dilemma as a simpler method of pre-selection induces a greater risk of overlooking good or even optimal alternatives.

4.8.2 Restrictions and aspiration levels

Alternatives are often subject to some restrictions. A department head, for instance, is only allowed to purchase items up to a predetermined limit. When filling a specific vacancy, no men would be considered if the proportion of women has to be increased. In practice, such restrictions are sometimes also referred to as "knock-out criteria" - alternatives that do not meet them are immediately "out". The advantage of such restrictions is their simplicity. They can help restrict the set of alternatives drastically, even if little is known about the set of alternatives. However, the risk of hastily eliminating good alternatives is correspondingly increased.

Almost the same applies to the aspiration level of an objective variable; these are quantitative maximum or minimum limits. With regard to investment decisions, a maximum payback period, for example, is often defined as a pre-selection criterion. Projects will only be considered if they guarantee that the invested amount will be recouped within, say, a maximum of 3 years. Or when filling a vacancy, work experience and business-related knowledge are required from the applicant. The grade point average is used as a (proxy) attribute for knowledge. Aspiration levels are determined as follows: candidates who do not have at least 5 years of work experience and gained their degree with a grade point average of at least 3.0 will not be considered further. In the case of uncertain expectations, the aspiration levels have to be determined probabilistically, e.g.: an investment opportunity must have at least a 90% probability of recouping the initial investment within the next 3 years.

One problem with aspiration levels is that certain objectives are treated as if their lack could not be compensated for. An applicant with a grade point average of 3.1 will be eliminated; he has no chance of making up the shortfall regarding the aspiration level with (good) attribute levels for other objectives, even if these are spectacular. Such a rule will (almost) never reflect the true preferences of the decision maker. A sensible use of aspiration levels should thus not be entirely mechanical. You should not eliminate an alternative that violates a certain aspiration level only marginally if it substantially surpasses your aspirations in many other dimensions.

However, the more essential problem with restrictions and aspiration levels is how to precisely define them. The more rigorous the requirements, the more easily it may happen that no alternative at all fulfills them. In that case, you have to relax the requirements - but which ones and by how much? Assume you are looking for a used car for less than €5,000, equipped with a CD player and automatic transmission and you cannot find one. Would you rather give up the CD player or the automatic transmission? Or would you prefer to pay an additional €500 rather than give up any of the two features?

The problem also exists in the other direction. If the requirements have not been high enough to reduce the set of alternatives significantly, they have to be tightened – but again, which ones and by how much? These questions can only be answered satisfactorily if the decision maker's preferences have been modeled and if the properties of all the alternatives are known. If this is the case, however, we do not need a pre-selection.

Restrictions and aspiration levels can be given exogenously, e.g. by rules, statutes or directives from your supervisor. In this case, they are not a problem within the given decision context. By contrast, it becomes a problem if the decision maker has to define them himself, purely for the purposes of reducing the set of alternatives.

Restrictions and aspiration levels should hence be used with caution. Specifying and modifying them is by no means a formalized procedure that can be tested for rationality. However, restrictions and aspiration levels may be necessary for reducing the number of alternatives. In such cases, they should be used in a way that minimizes the risk of prematurely discarding the best alternative.

4.8.3 Dominance

In the case of multiple objectives and in some circumstances, the alternatives can be assessed using the concept of dominance. In general, dominance refers to a situation in which one alternative can be identified as superior to another even though the information that is usually needed for evaluation is incomplete. Different concepts of dominance are applicable, depending on the type and specification of the preference model. At this point, we confine ourselves to the case of comparing alternatives under certainty and multiple objectives. In the previously presented example (Table 4-1), we have already identified alternatives that were dominated.

Every new alternative is compared pair wise with each of the existing alternatives. It holds that an alternative is *dominated* by another, if it is better than the other in no attribute but worse with respect to at least one attribute. The decision maker then does not need to care about the specific form of the additive value function. In particular, the weighting of objectives is of no relevance, alternatives that are dominated can be discarded.

This test for dominance is only applicable under the following conditions:

1. The system of objectives is known; all objective variables have been identified.
2. For each objective variable, the direction of preference is known. For example, maximization, minimization or a certain value of the objective variable are aimed for (e.g. the optimal average temperature at a vacation spot). This condition is violated, say, if one and the same attribute is used for various objectives with different directions of preference, like the size of an apartment as a measure of living comfort (maximization) and maintenance efforts (minimization). In this case, no dominance relationship between two differently sized apartments can be given.
3. There is a preferential independence among the objective variables so that an additive valuation model is appropriate.
4. The decision maker is interested only in the best alternative, not in the second best or in finding a complete order of all the alternatives. When applying the dominance test, it may easily happen that alternatives are eliminated that are slightly worse than the best one, but at the same time, better than some of the remaining alternatives.

Let's consider an example: a bank has trained five apprentices but can only offer a permanent job to two of them. It was decided to rely on the criteria "grade point average in high school", "grades during apprenticeship" and "personal impression" (also measured on a school grade scale). Here is a summary of the trainees' performance:

Trainee	High school GPA	Apprenticeship GPA	Personal impression
A	1.3	1.3	1.7
B	1.8	1.7	2.0
C	2.0	2.3	1.3
D	2.3	3.0	2.7
E	2.3	3.3	3.0

Trainee A dominates B, D and E. Since the bank manager has two positions to fill, he decides to take, in addition to A, the only non-dominated trainee C. However, this is not necessarily the best choice. B is not dominated by C and could turn out to be a better candidate once the manager's preferences are specified in detail.

Under the above-mentioned conditions, dominance is a sensible selection criterion; however, it requires a high level of information. If the system of objectives is still incomplete – e.g. the decision maker considers it possible to add some more objectives in the course of the decision process – dominance can only be tested for the already identified objectives. The same is true if the levels of some objective variables are not yet known. When looking at the applications for a vacant position, the attribute levels for education, grades, work experience, etc. can be determined. Other attributes such as personal impression or foreign language skills can only be assessed after a job interview. In this phase, dominance can be determined only partially, i.e. with respect to a subset of objectives; no alternative can ultimately be discarded at that point. (In the present example, you would not return the documents to any of the applicants hastily.)

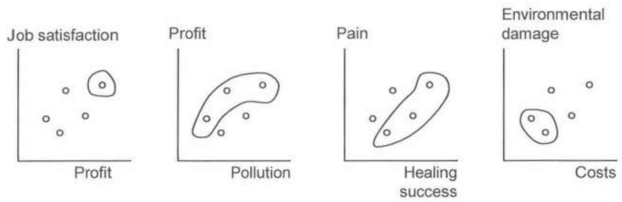


Figure 4-4: The marked alternatives are not dominated

Dominance relationships can be identified more effectively by graphical illustration than by just comparing numbers. In fig. 4-4, those alternatives are marked that are not dominated with respect to two attributes. Fig. 4-5 demonstrates that dominance relationships can also be illustrated graphically if there are more than two attributes; the scales, however, must have the same direction of preference. In the example, a car buyer has matched the levels of four attributes to four car models. For each attribute, a high level is preferred over a low level. Connecting the points results in an attribute profile for each automobile. An alternative dominates another if its profile lies above that of the other alternative. In the example, "Celia" dominates "Tornado" because it is better (or is at least not worse) with respect to all the relevant attributes.

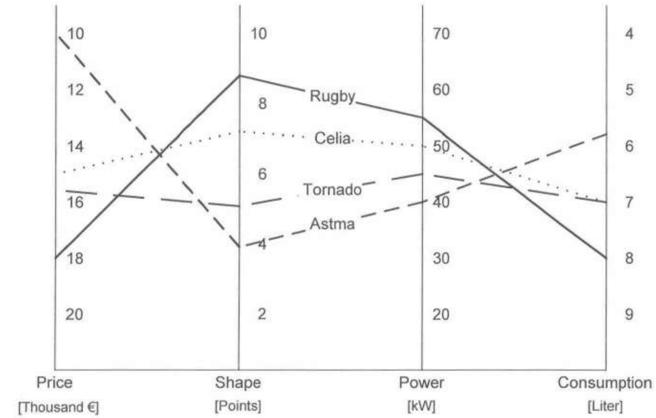


Figure 4-5: Attribute profiles of four car models

Questions and exercises

4.1

Your brother-in-law, Calle Noni, (the Italian restaurant owner from Exercise 2.5) is complaining about the bad profit situation. He has two goals: to increase the profits and still maintain the character of a plain, nice restaurant. Display the measures you consider appropriate in a means-end network.

4.2

Dr Sacrilege Ltd produces three chemicals denoted by *X*, *Y*, and *Z*. The business situation is bad and suspension of staff will be necessary. The production plan has to follow two goals: maximization of contribution margin and minimization of the number of employees laid off. The three products show the following properties concerning the attributes:

	X	*Y*	*Z*
Contribution margin in € per ton	95	75	50
Number of employees necessary per ton per month	0.20	0.25	0.35

The market capacity each month is 300 tons for *X*, 500 tons for *Y*, and 400 tons for *Z*. The production capacity for all three products together is 1,000 tons per month. Generate several production programs, none of them dominated with respect to the two goals.

4.3

The number of traffic accidents is too high. What can be done to reduce the number? First think about appropriate measures. In doing so, you will discover that besides achieving the objective of a reduction in the number of accidents, other fundamental objectives have to be considered as well. Create a means end network in which the fundamental objectives, the means to achieve them, and positive and negative effects you assume are depicted. Which additional measures will result from an extension of the problem to "reduction of traffic"?

4.4

Nibbleplants Inc. wants to extend its product line of chocolates with a chocolate bar that differs considerably from the competitors' products and that should reach a target group that is as big as possible. Try to develop this promising chocolate bar by decomposing the tasks.

4.5

The board of a German pharmaceutical group is discussing the US launch of a new drug against cardiac irregularities. The yearly sales in America will amount to 5 million packages with a probability of 30%, and 10 million packages with a probability of 70% at a price of €3 (converted from $) per package. Before the product can be launched on the American market, it has to be approved by the local health authorities. The probability of gaining approval is 70%; the approval procedure costs one million Euros.

The American market can be supplied either by opening a secondary plant in the U.S., exporting from Germany, or allocating the production to a local licensed party. The fixed costs of a secondary plant are €6 million per year and the variable costs per package produced are €0.80. The additional fixed costs of production in Germany are only €3 million, but the variable costs are €1.20 per package. A licensing fee of €0.50 per package could be obtained in the case of granting a license. The probability of finding a license holder is 50%; in order to find a license holder, an American consultancy is engaged, which has to be paid €0.5 million. Depict the possible multi-level alternatives by using a decision tree.

4.6

The set of alternatives stated in the table below is given; a higher level is always better. First, define the alternatives that are not dominated. For objective A, the minimum level is 55, and for objective B the minimum level is 26. Derive the efficient alternatives which comply with these requirements. Are there reasons why these minimum levels should not be used as requirements?

Alternative	Objective A	Objective B	Objective C
1	48	8	26
2	41	1	12
3	50	20	25
4	32	2	18
5	31	42	30
6	7	15	20
7	51	10	23
8	58	25	4
9	56	41	18
10	51	46	15
11	49	58	14
12	50	55	24
13	6	16	30
14	53	49	9
15	14	55	6
16	60	1	2
17	22	48	2
18	64	16	26
19	57	16	6
20	7	51	13

4.7

You're considering going on vacation to an island in the Mediterranean with a friend of yours after you've both completed the last exam of your studies. As a reward for the efforts you made during your studies, you plan to take a recreational dive course. Besides that, you are also interested in the island's culture. You define four objectives, namely the costs of the trip (costs), the diversity of the nightlife (party), the quality of the dive sites (diving), and the culture of the island (culture). Four islands are considered with the following attribute bundles. The attribute level represents the ordinal rank within the four alternatives. Identical levels yield identical ranks; a lower number indicates a better level (rank 1 is the best): Island A (costs 3; party 3; diving 2; culture 1), Island B (costs 4; party 3; diving 4; culture 3), Island C (costs 1; party 1; diving 3; culture 2), Island D (costs 1; party 1; diving 3; culture 2). Sketch the attribute profile for this decision situation. Will each alternative remain on the short list?

Case Study: Mexico City Airport

Source: Keeney and Raiffa (1976)u, Chapter 8.

Problem

Due to increasing air traffic in the late 1960ies, the Mexican Government was forced to think about measures to increase the capacity of the Mexico City airport. Two reports concluded two completely opposite recommendations. One of them recommended an extensive extension of the Texcoco airport, which was located close to the city; the other advised to shift the existing commercial air traffic to a yet-to-be-built airport close to the village of Zumpango, some 25 miles north of the city.

As the city had grown dramatically and as the existing airport was located in a densely populated area, an extension of the airport would carry substantial disadvantages in terms of noise pollution, risk of accidents and necessary relocation of residents for building new runways. On the other hand, the new airport would have the disadvantage of being far away from the city center.

In the summer of 1971, the Mexican ministry of public works conducted another study for which Keeney and Raiffa were hired as experts. The study was expected to confirm the ministry's opinion that Zumpango would be the optimal choice.

The static model

The ministry had given two main alternatives: extending Texcoco and constructing the new Zumpango. For the new airport, different sub-alternatives had to be distinguished since not all categories of air traffic necessarily had to be relocated to Zumpango. Four categories were chosen: (I) international flights, (N) national flights, (A) general air traffic, and (M) military. Since each category was meant to be located at only one airport, 16 different alternatives resulted.

Keeney and Raiffa further differentiated over time to consider changes in the type of usage. They defined 1975, 1985, and 1995 as points of decision. For instance, in 1975, Zumpango could be opened for the general air traffic, the international traffic could be relocated there in 1985, and from 1995 onwards, all air traffic could be operated from Zumpango. By this means, $16^3 = 4{,}096$ theoretical alternatives were derived. Most of them could be eliminated, however. Given that the military share amounted to less than 5%, alternatives that only differed with respect to this aspect could be aggregated. Other alternatives obviously didn't make sense, e.g. all combinations of relocating air traffic back and forth between the two airports. Overall, about 100 alternatives remained on the short-list.

After long discussions, the following six objectives together with the corresponding attributes were chosen by the ministry:

	Objective	Attribute
1	Minimal costs of construction and maintenance	Discounted costs in Million Pesos
2	Suitable capacity	Number of aircraft movement per hour
3	Minimal transfer time to the airport	Transfer time in minutes, weighted by the number of passengers from different districts of Mexico City
4	Maximal safety of the system	Number of fatal casualties or severely injured people per aircraft accident
5	Minimal social friction losses	Number of residents to be relocated
6	Minimal noise pollution by air traffic	Number of people exposed to a high level of noise pollution*

*90 CNR or more. This measure (Composite Noise Rating) is composed of the noise in decibel and the frequency of occurrence.

The next step comprised the estimation of probability distributions for the attributes' levels for each of the three decades considered. Furthermore, a multiattribute value function needed to be determined to value the different alternatives, details of which are omitted here. Two groups of alternatives came out best and were valued as approximately similar: (1) alternatives in which Zumpango would be built and would take over the I- and the N-traffic immediately, and (2) alternatives in which Zumpango would be built stepwise, starting with the I- or the N-traffic and later taking over the other category in either 1985 or 1995.

The dynamic model

In a second analysis, the experts accounted for the fact that it might be sensible to only derive a recommendation for the present (1971), given the high degree of uncertainty about future prospects. Five years later, a consecutive decision could be taken incorporating new developments. Possible aircraft accidents, changing demand structures, technical innovations – i.e. less noisy aircraft –, a changed attitude towards environmental issues as well as political and social changes were considered as uncertain factors. The intention was not to come up with an optimal multi-stage decision, however; the framework was formally narrowed to the first decision. However, as there would certainly be at least some impact of the first decision on later ones, the alternatives for 1971 hat to be valued again, additionally considering their level of future flexibility.

The alternatives for the first decision were derived by defining four effort levels for each airport, resulting in 16 combinations. The following figure displays the combinations:

Investment in Zumpango \ Investment in Texcoco	Minimum	Low	Medium	High
Minimum	1	2	3	4
Low	5	6	7	8
Medium	9	10	11	12
High	13	14	15	16

In the following, the different effort levels are explained.

Texcoco:

- Minimum: Ensure maintenance and procure safety gear.
- Low: Additionally extend runways; improve support buildings (e.g. terminals).
- Medium: Additionally acquire and prepare land for new runways and passenger facilities.
- High: Built a completely new airport Texcoco.

Zumpango:

- Minimum: At most, acquire the land.
- Low: Acquire the land, build a runway and plain passenger facilities.
- Medium: Acquire the land, build the first runway and plan the second, install larger passenger facilities and build a feeder street for the highway toward Mexico City.
- High: Build a couple of runways, large passenger facilities, comprehensive connection to the road network; hence building a new major airport in Zumpango.

Compared to the static model, the system of objectives was extended. The six objectives from the static model were subsumed under the main objective "effectiveness". The "flexibility" which an alternative would have five years onwards (from 1971) was added as another objective. Furthermore, "political consequences" for the president as the highest-ranked decision maker as well as the administration involved were modeled as an objective. Finally, "external consequences" representing the impacts on, for instance, regional development or national prestige, were considered.

The number of 16 alternatives was reduced by preselecting some alternatives after intense discussion between the experts and the administration. Alternative 1

was ruled out because it did not fulfill the requirements for the immediate future. Alternatives 7, 8, 11, 12, 15, and 16 were eliminated as high investments in Texcoco would render parallel investments in Zumpango unnecessary. Alternatives 3 and 4 were merged as they hardly differed from each other. Eight alternatives remained.

The alternatives' valuations regarding their effectiveness were adopted from the static model. Members of the ministry valued the alternatives in terms of the additionally defined objective variables. This was done using an ordinal scale, i.e. the alternatives were given ranks; the best one received the rank 1, etc. Alternatives judged as equal were given identical ranks. The following table shows the results:

Alternative	Effectiveness	Flexibility	Political Consequences	External Consequences
2	7	1	3	3
4*	8	7	4	7
5	3	2	3	1
6	1	3	2	3
9*	4	4	5	2
10	5	5	1	4
13*	2	6	7	5
14*	6	8	6	6

The asterisk denotes the dominated alternatives. For example, alternative 4 turned out to be worse than alternative 2 in every aspect. The dominated alternatives were eliminated, but not without Keeney and Raiffa playing devil's advocate to ensure that nothing in favor of these options was missed. Only 2, 5, 6, and 10 remained on the short list.

In the last step, the experts defined the remaining alternatives in more detail; no. 5 was split into two subsets:

2 In Zumpango, only acquire the land. In Texcoco, the two main runways and the ramp are extended; freight and parking facilities are built; no new passenger terminals.

5a In Zumpango, build a runway, a few terminals and a plain connection to the road network. Acquire enough land to build a new major airport. In Texcoco, conduct only maintenance and enhance safety measures.

5b Like 5a, except that in Zumpango the land necessary for the first development stage only is acquired.

6 In Texcoco, only one runway is extended, the other enhancements are as in alternative 2. In Zumpango, the land for a major airport is acquired and a runway as well as some passenger facilities and access roads are built.

10 In Texcoco, identical as in alternative 6. In Zumpango, build two runways, larger passenger facilities and access roads.

These five alternatives were reanalyzed and evaluated again. The new ranks are displayed in the following table:

Alternative	Effectiveness	Flexibility	Political Consequences	External Consequences
2	3	1	4	4
5a	3	2	3	3
5b*	4	4	5	5
6	1	3	1	1
10*	1	5	2	2

Again, some alternatives were dominated (*). Three alternatives remained. A methodological-formal decision was not conducted; alternative 6 was considered by the members of the ministry to be the best.

Consequences
The ministry of public works was surprised by the results. After expecting simply a confirmation for the option of immediately building Zumpango, the static model indicated that a stepwise approach could be just as valuable. Given a dynamic view in combination with the additionally considered objective variables of flexibility as well as political and external consequences, the alternative initially preferred turned out to be completely dominated. As a consequence, the ministry opted for a more flexible attitude and at the end of 1971 recommended to the president a stepwise development of Zumpango.

Chapter 5: Decision making under certainty with one objective

5.0 Summary

1. Completeness and transitivity are rationality demands which we postulate for a preference when we consider decision making under certainty with one objective.
2. If a preference is a complete and transitive ordering, it is always possible to find a value function that represents that preference.
3. There are measurable and non-measurable value functions; different from non-measurable value functions, measurable value functions allow conclusions to be drawn about the strength of the preference.
4. You will learn three methods of determining the value function: the direct-rating method, the difference standard sequence technique and the bisection method.
5. Consistency checks are an essential element of the process of determining the value function.

5.1 Value function and preference

In this chapter, we wish to return to the concept of the decision maker's preference regarding a single objective. In doing so, we assume every alternative to have exactly one consequence, i.e. the outcome is certain. By means of this simple decision situation, we explain how decision makers' preferences can be displayed via value functions and what techniques can be used to determine value functions. While reading this chapter, you might occasionally wonder why such complicated methods are applied to seemingly trivial problems. Why, for example, do we need to determine a value function for a university graduate choosing his first job when we have decided to assume that he is only interested in salary and if the most attractive (highest paid) job option can simply be picked from a list? It is not because there appear to be many relevant decisions under certainty and with only one objective in daily life which require the use of these tools and concepts discussed here; indeed the value function concept is largely redundant under such conditions. We initially present this simple example because a thorough understanding of the theory presented in this chapter provides a very useful basis for the following chapters. For the simple case of decision making under certainty with one objective, a procedure for displaying and measuring preferences is presented here, which we will subsequently transfer to cases of multiple objectives, multiple states, and multiple periods – and hence much more difficult decision problems.

In this chapter, we assume as given a finite set of alternatives denominated $A = \{a, b, c, ...\}$. We refer to the consequence of an alternative a with respect to

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_5, © Springer-Verlag Berlin Heidelberg 2010

the relevant objective, or – to be more specific – with respect to the chosen attribute X, as x_a. This consequence is also called "level of target achievement" or "attribute level of the alternative". The attribute levels can be given on different types of scales (nominal, ordinal or cardinal).

As an example of decision making under certainty with one objective, let us consider the decision problem of a business student that has recently graduated from university and has to choose his first job. In the course of this chapter, we assume that the decision maker considers his annual salary to be the only objective; the consequences of the alternatives are hence measured in € per year. The decision maker takes the job alternatives displayed in Table 5-1 into consideration:

Table 5-1: Three job offers with one attribute

	Alternative	Consequence
(*a*)	Business Consultancy	€80,000
(*b*)	University	€50,000
(*c*)	Sailing Instructor	€30,000

In the following procedure, we would like to determine a decision maker's preference and represent it by means of a function. In this example, of course, this would not be necessary in order to determine the optimal alternative. To you (and to our student), €80,000 certainly has a higher value than €50,000 or €30,000. However, as we mentioned before, the ability to represent and measure preferences is an important precondition for solving more complex decision problems. The procedures used to support decision making with multiple objectives (Chapter 6) as well as intertemporal decision making (Chapter 11) assume that you are able to suitably represent preferences regarding a single objective. Additionally, when discussing decision making under uncertainty (Chapters 9 and 10), many of the concepts presented in this chapter will be used again.

A decision maker has a *preference* between two alternatives $a, b \in A$ if he either prefers a to b ($a \succ b$) or b to a ($b \succ a$) or if he is indifferent between a and b ($a \sim b$). Within the prescriptive approach of decision analysis, we wish to postulate some rationality demands on the decision maker's preferences.

In general, we require the preference relation (in short "the preference") to be a complete and transitive ordering. A preference is *complete* if the decision maker has a preference for any pair of alternatives. It is *transitive* if for any three alternatives a, b, and c, the following holds: from $a \succ b$ and $b \succ c$ follows $a \succ c$. A decision maker who claims in the given situation to be unable to compare the alternative "Sailing Instructor" to the alternative "Business Consultancy" is considered to be just as irrational as a decision maker who prefers "Sailing Instructor" to "Business Consultancy", the latter to "University" and "University" to "Sailing Instructor".

In the example, we derived the decision maker's preference from an evaluation of the consequences. Generally, in case of expectations under certainty, the function that evaluates the alternatives based on its consequences and that reflects the preference, is called the *value function*. The value function is a mathematical representation of the preference. Let it be denominated by v and defined formally as follows.

Definition 5.1 (Value function)
A value function v is a function that assigns a real number to each alternative, such that the value of alternative a is greater than the value of alternative b if and only if the decision maker prefers alternative a to alternative b:

$$v(a) > v(b) \Leftrightarrow a \succ b$$
$$a, b \in A \quad \text{(analogous for } \sim, \prec\text{)}.$$

Not every imaginable preference can be represented by a value function. Preferences which do not fulfill the transitivity axiom cannot be mapped via value functions as we have defined them here. Since values are real numbers and ">" is a transitive ordering of the real numbers, the preferences have to be transitive as well. It is similarly obvious that preferences have to be complete to allow a representation by means of a value function. The simple reason is that the size of any two real numbers can be compared and thus this also has to hold for alternatives. Fortunately there is not much more than completeness and transitivity as preconditions for the existence of a value function to represent a preference. It holds:

Proposition 5.1: If a preference is a complete and transitive ordering (and the technical condition holds that the set A is countable, see Krantz et al. 1971, p. 39), a value function will always exist that represents the preference.

Since we have taken the axioms "completeness" and "transitivity" as a basis for rational decision making, we can assume in the following that preferences can be represented by value functions (for exceptions caused by a violation of the technical conditions, see Rauhut et al. 1979, p. 34 f.) Through this insight, a great deal has already been achieved: you now "only" need to learn how the decision maker's value function can be determined in order to know his preference for arbitrary alternatives and to be able to identify the optimal alternative.

The representation of a preference by a value function might seem trivial at first glance, almost too cumbersome for something this simple. However, the concept is based on the following assumptions and basic considerations that we will also have to think about when mapping much more complex preferences over the course of the book:

- We defined the preference for decision making under certainty with one objective. For this decision problem, conditions were attached to the preference from which the existence of a value function can be derived.
- A rational preference under certain expectations should fulfill the axioms of "completeness" and "transitivity". When considering decision making under certainty with multiple objectives or decision making under uncertainty, the respective axioms allow for an easy mathematical representation of the preference.

- The more complex decision situations that we present in the following chapters require the preference to meet the axioms of completeness and transitivity as well. Nevertheless, decision makers often violate these axioms when making decisions intuitively. In the very first chapter of this book, we have already pointed out the possible discrepancy between actual decision behavior (subject of descriptive theory) and rational decision behavior (subject of decision analysis, i.e. prescriptive theory). Therefore, at this point, we have to discuss whether it is legitimate to use the axioms of transitivity and completeness as the basis for rational decision behavior.
- Transitivity is regarded as the basis for rational behavior by many authors in the field. From the intransitive behavior $a \succ b$, $b \succ c$, and $c \succ a$, a so-called "money pump" can be constructed. To illustrate the point, we assume the decision maker to be equipped with alternative a. If he prefers c to a, he will at least pay a small amount of money α to exchange a for c. Accordingly, he will pay an amount α' in order to exchange b for c and an amount α'' in order to exchange a for b. Considering these three exchanges together, he paid a total amount of $\alpha + \alpha' + \alpha''$ and yet again received his initial alternative. Under these circumstances, the decision maker will not stick to his intransitive preference for long. Of course, the mechanism of the "money pump" itself is again based on specific preconditions; for instance, there has to be an exchange partner at any time. Nevertheless, it offers a strong argument for considering intransitivity to be irrational.
- Mainly in the French decision making literature, a preference theory is dealt with that considers the categories "indifference", "weak preference", "strong preference" and "incomparableness" (Roy 1980). Alternatives could be incomparable, for instance, for two unknown dreadful alternatives, e.g. "would you prefer to die in a car accident or a plane crash"? However, we hope that we will be confronted with "better" alternatives and will focus primarily on the prescriptive aspect. A rational decision maker should be able to decide for each pair of alternatives which alternative he prefers or whether he is indifferent between them.
- The value function has been defined formally as a function that represents the decision maker's preference. An interpretation of the value term with regard to its content has not been conducted.

A preference can usually be mapped to more than one value function. If a second value function v' is derived from the original value function v via a strictly monotonically increasing transformation, v' sorts the alternatives exactly the same way as v does, i.e. it holds that:

$$v(a) > v(b) \Leftrightarrow v'(a) > v'(b). \tag{5.1}$$

Both value functions are hence equivalent as they represent mappings of the same underlying preference.

We can explain the connection between preference and equivalent value functions with an example again. Let the business student have the following preference with respect to his job alternatives:

$$\text{Consultancy} \succ \text{University} \succ \text{Sailing Instructor}.$$

Table 5-2: Equivalent value functions

	Consultancy	University	Sailing Instructor
v	10	8	7
v'	20	16	14
v''	20	16	8

A value function must assign real numbers to these three alternatives in such a way that the evaluation of "Consultancy" is higher than the evaluation of "University", which itself has to be higher than the evaluation of "Sailing Instructor". The value function v in Table 5-2, which assigns the real numbers 10, 8 and 7 to the alternatives, fulfills this condition. The second value function v' was derived from v by multiplying all values by the factor 2. Multiplication by a positive number is a strictly monotonic transformation, and accordingly the value function v' sorts the alternatives exactly as does the function v. The strictly monotonic transformation of the function v into function v'' results in an equivalent value function as well.

The type of value function that has been defined so far maps the *order* of alternatives. It does not contain any information about the *strength of preference* between the alternatives. Therefore, this value function is sometimes referred to as an *ordinal* value function. Statements like "the value difference between University and Sailing Instructor is smaller than the value difference between Consultancy and University" (with respect to value function v) have no validity for the concept of ordinal value functions. This becomes evident when considering the fact that the value differences do not remain the same when conducting the feasible transformation of the value function from v to v''.

As opposed to the concept of ordinal, non-measurable value functions, the concept of *measurable* value functions also allows for mapping the strength of preference. To derive the concept of measurable value functions, let the decision maker again have a complete and transitive preference regarding the alternatives. In addition, he must have a preference with respect to the transitions between alternatives. The transition from alternative a to alternative b is noted as "$(a \rightarrow b)$". The transition from sailing instructor to working at the university (Sailing Instructor$\rightarrow$University) is characterized by the fact that working at the university adds an additional annual salary of €20,000 as compared to the annual salary of the sailing instructor. Note that not the final state of the transition is considered but rather the perceived difference in value caused by the transition from one alternative to another.

In order to make statements about the strength of preference, the decision maker must be able to compare different transitions with one another, i.e. he must have

clearly ordered preferences with respect to these transitions. For arbitrary alternatives a, b, c or d, it must hold that $(a \rightarrow b) \succ (c \rightarrow d)$ (or $\sim$ or $\prec$). In the example, this is a statement of the type: I prefer the improvement from sailing instructor to assistant professor over the improvement from assistant professor to consultant (Sailing Instructor $\rightarrow$ University) $\succ$ (University $\rightarrow$ Consultancy). It is assumed in the following that this preference regarding transitions is also complete and transitive. The measurable value function can then be defined as follows.

Definition 5.2 (Measurable value function)
A value function v is called *measurable* if it is a value function according to definition 5.1 with the additional property that the transition from alternative a to alternative b is preferred to that from alternative c to alternative d if and only if the value difference between b and a is greater than the value difference between d and c:

$$v(b)-v(a) > v(d)-v(c) \Leftrightarrow (a \rightarrow b) \succ (c \rightarrow d)$$

$$\text{with } a, b, c, d \in A \quad (\text{analogous for } \sim, \prec).$$

The expression *measurable value function* can lead to misunderstandings. Both ordinal as well as measurable (i.e. cardinal value functions) and even utility functions (which will be defined in Chapter 9) *all* reflect the decision maker's preference based on theoretical considerations regarding measurability. The distinct significance of these functions results from the characteristics of the preference that they "measure".

As you have seen, the measurable value function is measured on a cardinal scale. Aside from the ordinal comparison, it is also possible to interpret differences in the v-values. Formally, aside from the generation of differences, an addition of v-values also seems feasible. However, since only those characteristics of the preference that are related to the generation of differences were required when defining the representative value function, adding v-values does not have any empirical meaning.

Furthermore, the division and multiplication of two v-values is not reasonable either, i.e. statements of the form "alternative a is preferred three times as much as alternative b" are not compatible with the concept of measurable value functions. For general theoretical basics of measurability see Krantz et al. (1971) and Roberts (1979).

Analogous to the ordinal (non-measurable) value function, it holds that:
Proposition 5.2: A measurable value function (representing the preference) exists if and only if the preferences for alternatives as well as the preference for transitions between alternatives are complete and transitive orderings (and some further technical conditions hold; see Krantz et al. 1971).

For measurable value functions it also holds true that more than one function exists that represent the preference. In fact, measurable value functions are unique up to positive and linear transformations. If v is a measurable value function,

$v' = \alpha \cdot v + \beta$ with $\alpha > 0$ is a measurable value function for the same preference as well. Again, we can illustrate this with an example:

Table 5-3: Equivalent measurable value functions

	Consultancy	University	Sailing Instructor
v	10	8	7
$v' = 2\,v$	20	16	14
$v'' = v' - 10$	10	6	4

It holds: (University $\rightarrow$ Consultancy) $\succ$ (Sailing Instructor $\rightarrow$ University). In the example in Table 5-3, the measurable value function v is given and represents the preference "consultancy is better than university is better than sailing instructor" and the preference "the transition from university to consultancy is better than the transition from sailing instructor to university". The value functions v' und v'' represent the same preference. All equivalent value functions represent exactly the same preference (more precisely, a preference for alternatives and one for transitions between alternatives). The measurable value function reflects "only" these two types of preference – no more and no less.

The differentiation between value functions and measurable value functions might seem a bit academic at this point. However, whether or not a value function is measurable has important implications, not only for the method used to determine a value function but also later on when we discuss the trade-off between value differences in multiattribute decision making.

5.2 Methods for determining value functions

5.2.1 Introduction

The reliable determination of value functions is an important step in the decision making process. Once the value function of a decision maker has been determined, the preference for a given set of alternatives can be derived directly. Because the value function is intended to represent the decision maker's preference, it has to be determined based on preferences. The decision maker is confronted with simple paired comparisons or is asked to construct indifference statements for simple alternatives; the value function is then derived from these statements.

The value function is plotted in a diagram which shows on the horizontal axis the attribute levels of the relevant objective variable X that can be assumed by the given or conceivable alternatives. Let x^- be the minimum attribute level and x^+ the maximum attribute level. The values of the function are plotted on the vertical axis. The value function is usually normalized on the interval [0, 1]. Figure 5-1 presents a possible value function for the objective annual salary in a range from €30,000 to €80,000. If the depicted value function can be interpreted as a measurable one, it becomes apparent that the decision maker values a transition from €30,000 to €50,000 exactly as much as a transition from €50,000 to €80,000.

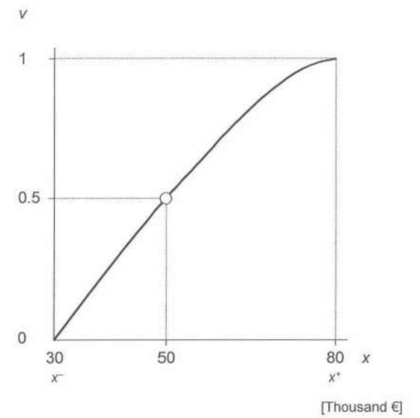


Figure 5-1: Possible value function for annual earnings between €30,000 and €80,000

We have already pointed out that the value function must be determined based on preferences. It is therefore not sufficient to present different types of value functions (e.g. concave, monotonically increasing or convex, monotonically increasing functions) to the decision maker and ask him to decide in favor of one these types. All the methods presented below ask the decision maker to compare attribute levels and to evaluate them in a clearly specified manner.

A preference with respect to an attribute is usually not permanently stored in the decision maker's mind, i.e. it is not available at all times; it needs to be elicited through a skillful questioning technique. In the process of questioning the decision maker – which can also be self-questioning – errors can occur, making *consistency checks* essential when determining value functions.

The literature offers an abundance of methods for determining value functions (see von Winterfeldt and Edwards 1986 as well as Farquhar and Keller 1989). Below, we present three methods and initially assume that the value functions are monotonically increasing, i.e. the decision maker prefers a higher attribute level to a lower one. The methods apply analogously to monotonically decreasing value functions. Additionally, the attribute is initially assumed to be continuously scaled.

All the techniques that we present determine only a few points of the value function. When considering continuous attribute levels, the levels of consequences which have not been elicited directly can be determined by various methods. On the basis of the elicited points, a suitable curve can be fitted with the aid of mathematical techniques. In order to do so, the general shape of a function is chosen that approximately fits the elicited points (e.g. an exponential or quadratic function). The parameters of the general function are determined by the mathematical method in such a way that the curve best fits the elicited points. Current decision-support software usually offers such interpolation procedures. Alternatively, the value function can simply be interpolated between the given points, either by a piecewise linear function or by some other function type. Figure 5-2 illustrates the two approaches. In Figure 5-2a, a function (from some family of functions) is chosen that best fits the elicited points. The value function in Figure 5-2b is derived through piecewise linear interpolation.

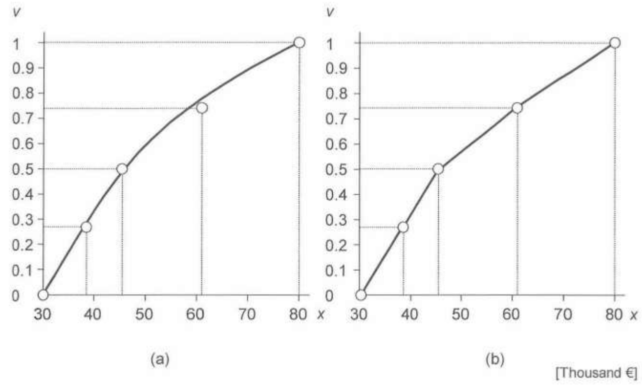


Figure 5-2: Possibilities for obtaining a value function from single points

5.2.2 The direct-rating method

The direct-rating method is the easiest way to determine the value function. Each alternative (i.e. each attribute level or consequence) is evaluated directly. First, the best and the worst parameter values are specified. Second, the order of alternatives according to the decision maker's preferences regarding the respective attribute levels is determined. Third, a direct evaluation of the alternatives is conducted. The alternative with the best consequence receives 100 points and the worst receives 0 points. The decision maker has to assign points between 0 and 100 to the intermediate alternatives in such a way that the ranking order of the alternatives is represented. For measurable value functions, the assignment of points has to be conducted in such a way that it also reflects the order of preference strength differences. After the value function has been normalized to the interval [0, 1], it can then be plotted in a fourth step.

If there are many alternatives that are to be judged, it is usually not possible to evaluate each alternative directly. The decision maker will conduct the procedure described above with a few alternatives whose consequences are distributed relatively evenly in the interval of possible attribute levels and will then fit an adequate curve to the elicited points.

Finally, the value function that has been determined so far has to be tested for consistency. Generally speaking, a consistency check compares preference statements that are deduced from known parts of the value function with preference statements not yet considered. We will return to a discussion of how to cope with inconsistencies after we have presented all the methods (see Section 5.2.5).

We illustrate the direct-rating method with an example. Let the decision maker evaluate the alternative job options in Table 5-4. Again, let the annual salary be the only relevant objective.

Table 5-4: Five alternatives and their annual salaries

Alternative	Consequence
(*a*) University	€50,000€
(*b*) Banking Trainee	€55,000€
(*c*) Business Consultancy	€80,000€
(*d*) Sailing Instructor	€30,000€
(*e*) Management Assistant	€65,000€

First, the worst and the best consequence are determined: $x^- = €30{,}000$ and $x^+ = €80{,}000$.

Second, the alternatives are arranged according to the preference, with respect to the consequences: $c \succ e \succ b \succ a \succ d$.

Third, a direct evaluation of the alternatives is conducted. Let it be $c = 100$, $e = 90$, $b = 80$, $a = 70$, and $d = 0$ points.

Fourth, the value function has to be normalized and drawn. Figure 5-3 displays the value function of the example. The value function has been derived by piecewise linear interpolation.

Fifth, a simple consistency check verifies that the original order of the alternatives coincides with the ordering induced by the assigned points.

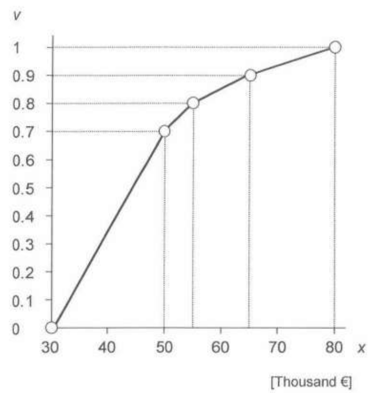


Figure 5-3: Value function for annual salary in the range from €30,000 to €80,000 elicited via direct-rating method

5.2.3 The difference standard sequence technique

A second method, based on the construction of equal value differences, is known in the literature as the "difference standard sequence technique". First, the worst relevant attribute level is determined. In the example of the job starter, the consequence of €30,000 produces the worst relevant level of the attribute "annual salary". The value of the worst attribute level is defined as zero, i.e. in the example v(€30,000) = 0.

Second, the range of possible consequences is examined and a unit is defined that represents approximately 1/5 of the total range. If the worst attribute level is increased by this unit, a second attribute level is received; this level is assigned a value of one. In the example, the chosen unit could be €10,000 and we thus set v (€30,000 + €10,000) = v(€40,000) = 1, i.e. we assign a value difference of one to the transition from €30,000 to €40,000.

Third, other transitions are measured by aid of the norm transition, i.e. in the example by aid of the transition from €30,000 to €40,000. We initially search for the attribute level z for which it holds that:

$$(€30{,}000 \rightarrow €40{,}000) \sim (€40.000 \rightarrow z).$$

Since v(€40,000) = 1 and v(€30,000) = 0, it must hold that $v(z) = 2$. Building on this information, the next level z' can be determined, such that the transition from €30,000 to €40,000 is just as attractive as that from z to z'. We assign a value 3 to the level z'. The variable z' can also be elicited via a comparison with the transition from €40,000 to z. We assign higher values to greater consequences until the new

consequence is equal to the best consequence or it outruns the range of possible consequences for the first time.

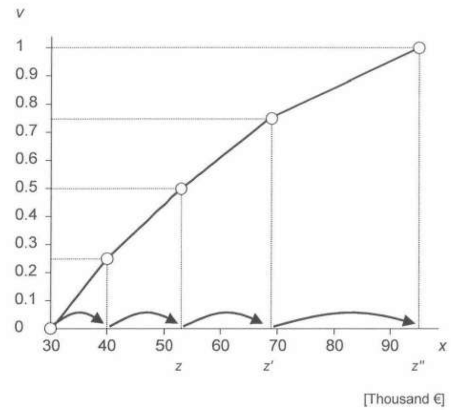


Figure 5-4: Determination of the value function by the difference standard sequence technique

Fourth, the values assigned to the consequences are normalized, i.e. divided by the highest value assigned to a consequence. We obtain v(€30,000) = 0, v(€40,000) = 0.25, v(€53,000) = 0.5 (for z = €53,000), v(€69,000) = 0.75 (for z' = €69,000) and v(€95,000) = 1. Afterwards, the value function can be drawn. Figure 5-4 depicts the process and the resulting value function for our example.

Fifth and finally, a consistency check has to be conducted. It would here be possible, for instance, to ask for the attribute level that is in the middle of the interval $[x^-, x^+]$ from a value perspective (see the bisection method in Section 5.2.4). This intermediate attribute level has to coincide with the consequence that has a value of 0.5. The difference standard sequence technique could also be repeated with a different starting unit, i.e. a different interval that is assigned a value difference of 1.

These consistency checks clarify the basic idea behind this important step: using a specific method, the decision maker is questioned in such a way that he needs to determine the value of an attribute level in a different manner. A consistent decision maker will assign a unique value to each consequence, independent of the chosen questioning technique.

The disadvantage of this method follows from the fact that the best consequence of the original attribute range does not have to coincide with the highest consequence resulting from the questioning technique. This is not a serious problem though. In this case, one will either stop with the questioning process early and extrapolate the value function beyond the last point or go beyond the highest consequence of the original range and determine the value function for a larger interval.

However, one has to be careful with respect to the normalization. This is particularly so if a value function is elicited by different methods in order to compare these functions for a consistency check. If a value function v was elicited by the difference standard sequence technique and normalized on a different (e.g. larger) attribute range, it cannot be directly compared to a value function v' that was normalized on the original attribute range. However, it is relatively easy in this case to rescale the already interpolated function v in such a way that it also assigns values between 0 and 1 to the original attribute range $[x^-, x^+]$.

5.2.4 The bisection method

The bisection method is also known as the "mid-value splitting technique". First, the worst and the best attribute levels are determined, as is the case with the direct-rating method. In the example, it holds that $x^- = €30{,}000$ and $x^+ = €80{,}000$.

Second, the decision maker is asked to specify the attribute level that bisects the interval $[x^-, x^+]$ from a value perspective. This consequence is called $x_{0.5}$. For this consequence, it holds:

$$(x^- \rightarrow x_{0.5}) \sim (x_{0.5} \rightarrow x^+).$$

Due to the measurability of the value function, this indifference is equal to $v(x_{0.5}) - v(x^-) = v(x^+) - v(x_{0.5})$. If one sets $v(x^-) = 0$ and $v(x^+) = 1$, it follows that $v(x_{0.5}) = 0.5$. In the example, assume that $x_{0.5} = €50{,}000$, i.e. the transition from €30,000 to €50,000 is as attractive as that from €50,000 to €80,000.

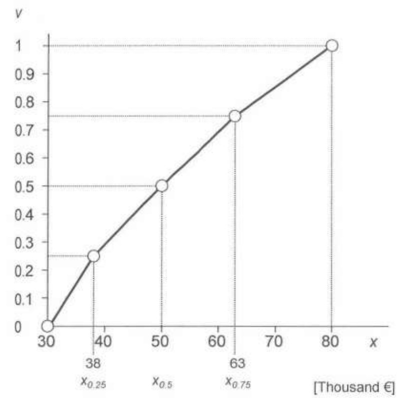


Figure 5-5: Determination of the value function according to the bisection method

Third, the bisecting attribute levels for the intervals $[x^-, x_{0.5}]$ and $[x_{0.5}, x^+]$ are determined analogously and denominated with $x_{0.25}$ and $x_{0.75}$, respectively. It holds that $v(x_{0.25}) = 0.25$ and $v(x_{0.75}) = 0.75$. In the example, let $x_{0.25} = €38{,}000$ and $x_{0.75} = €63{,}000$. The constructed intervals can be further subdivided in an analogous manner. However, five support points (x^-, $x_{0.25}$, $x_{0.5}$, $x_{0.75}$ and x^+) are often sufficient to determine and draw the value function.

Fourth, the value function is outlined in the relevant interval from €30,000 to €80,000. Figure 5-5 presents the value function determined in the example.

Finally, consistency checks also have to be conducted for the bisection method. The easiest consistency check is to ask the decision maker for the center of the interval $[x_{0.25}, x_{0.75}]$ from a value perspective. This consequence has to be equal to $x_{0.5}$. In the example, the decision maker has to be indifferent between a transition from €38,000 to €50,000 and a transition from €50,000 to €63,000.

5.2.5 Comparison of methods, consistency checks and non-monotonic value functions

Comparison of methods

The three methods presented in the last few sections differ more or less with respect to the mental strain that they cause for the respondent. At first glance, the direct-rating method seems to be the simplest of the three as it only asks the decision maker to assign points to consequences. However, when properly applied, it actually turns out to be the most mentally challenging of the procedures. In order to obtain a measurable value function, the respondent has to assign the points in a way that also reflects his strengths of preference. This is by no means a simple task and many respondents might not even be sure how to interpret the request to generate "point assignments that reflect preference strength". Strictly speaking, indeed the direct-rating method amounts to not much more than asking the decision maker to plot some support points of his value function (in such a way that value differences reflect preference strength). From this perspective, the direct-rating method provides the least support to the decision maker in his endeavor to determine his value function.

The difference standard sequence technique and the bisection method differ substantially from the direct-rating method. When applying both these methods, it is not necessary to interpret the abstract concept of "assigning points to reflect preference strength". The decision maker has to do nothing but state a preference with respect to some explicit transitions ("I would be indifferent between these two transitions"). For the difference standard sequence technique, the generation of indifference statements should even be somewhat easier than for the bisection method because only one transition is modified while the other is pre-specified; the bisection method forces the respondent to adjust both transitions simultaneously, making it a little more difficult for the decision maker to determine a reliable and robust indifference statement.

Consistency checks

A detailed examination of the inconsistencies which arise is an indispensable precondition for any correct representation of the preference. Inconsistencies can occur due to both unsystematic and systematic errors. The prescriptive approach of decision analysis accepts the fact that answers to difficult preference questions can be volatile. A sound decision support procedure has to incorporate the possible mental limitations of the decision maker. Inconsistencies can occur within a single method of elicitation; they can, however, also occur because value functions are obtained from different techniques.

If inconsistencies occur, the decision maker or decision analyst will start by analyzing the contradictory statements. The decision maker is asked to change previous preference statements in such a way that the inconsistency disappears. In the course of an interactive procedure, increasing consistency can be achieved by repeatedly correcting inconsistent statements until finally a unique value function is derived. If inconsistent valuations are considered to be driven by an unsystematic error, this error can at least be reduced (if not completely mitigated) by taking the mean of the valuations. In this case, the value function results from the average of the inconsistent valuations. Inconsistencies could also be interpreted as signaling that a decision maker does not have an exact, complete preference. In this case, the inconsistent statements define the possible range of the feasible class of value functions. We will elaborate on the concept of incomplete information regarding the preference in more detail in Section 5.3.

Non-monotonic value functions

We have explained how to determine value functions using an example in which a preference regarding salary was represented. The derived value functions were *monotonic* and the attribute values *continuous*. For value functions that are neither monotonically increasing nor decreasing, the best (or worst) consequence in the interval $[x^-, x^+]$ has to be determined first, independent of the method used. Afterwards, the interval can be split into two (or possibly more) subintervals on which the value function is monotonically increasing or decreasing. The techniques earlier presented can then be applied separately to both intervals. For example, think of the number of bananas (glasses of beer, wine, etc.) that you would like to consume on a single day. Let the interval of possible consequences range from zero to ten bananas. Certainly, there is an optimal value for you somewhere within that interval. The authors, for instance, like three bananas best and would thus have to determine one value function for the interval [0 bananas, 3 bananas] and one for the interval [3 bananas, 10 bananas]. It holds that $v(3 \text{ bananas}) = 1$. Depending on whether 10 bananas or 0 bananas is considered to be the worst attribute level, the value zero is assigned, say for instance $v(10 \text{ bananas}) = 0$. Thus, the value function on the interval [3 bananas, 10 bananas] is normalized to [0, 1]. In order to align the value function on the interval [0 bananas, 3 bananas], a further indifference statement is needed. It relates the consequence of 0 bananas to a suitable consequence in the interval [3 bananas, 10 bananas] and thereby specifies the value to

be assigned to the outcome of 0 bananas when the value function is rescaled on this interval.

Non-monotonic value functions often result from an inappropriate structuring of the system of objectives. Keeney (1981) argues that by using a thorough process (see Chapter 3), objectives can usually be found that lead to monotonic value functions. Objectives with non-monotonic value functions, in particular, can be split into monotonic lower-level objectives. In the example, the correct objective could be the satiation level (or satiation level and feeling of fullness when split into lower-level objectives).

If the attribute values are *not available in a continuous form*, neither the bisection method nor the difference standard sequence technique can be used to determine the value function. Discrete parameter values have to be evaluated directly. If the value function for the color of a car is to be determined and the possible attribute levels are yellow, white and pink, points have to be assigned directly to the attribute levels.

The techniques for determining value functions cannot be described in such a way that one is able to learn them theoretically and then apply them to every decision problem without any difficulties. A correct application of the techniques requires lots of practice with real problems. With regard to their subtleties, each method needs to be adjusted with regard to the decision maker and the decision problem in question. We are confident that you will put this to the test by practicing the techniques at the next opportunity and trying to determine the value functions of some friends with respect to the annual number of days off work, annual salaries, or even the number of jelly beans consumed per day.

5.3 Incomplete information

Up to now we have assumed in this chapter that the decision maker has a complete and transitive preference. However, there are situations in which the axiom of completeness can be violated. A decision maker may not have a formally complete preference or may not be willing to express his preference exactly. For example, when using the bisection method, he might be sure that from a value perspective, the center of the interval [€30,000, €80,000] is located somewhere between €50,000 and €55,000 but not be able to decide in favor of any exact value. If he is forced by the method to specify an exact value, he might reject the results derived from the value function. It is not possible to represent such an incomplete preference in the value calculations that have been introduced so far. However, the calculations can be extended quite easily in order to represent incomplete information about preferences. Instead of a single value function, a class of value functions compatible with the information provided by the decision maker is used. Let I denote the information at hand, and let $V(I)$ denote the class of value functions that is compatible with this information. All value functions of the class $V(I)$ are classified as feasible. One alternative is preferred to a second alternative, if the value of the first alternative is at least as great as the value of the second alternative for all feasible value functions and higher for at least one value function. Formally, this can be defined as:

$$a \succeq_{V(I)} b \Leftrightarrow v(a) \geq v(b) \text{ for all } v \in V(I)$$
$$\text{and } v(a) > v(b) \text{ for at least one } v \in V(I)$$
$$a, b \in A \quad \text{(analogous for } \sim, \prec\text{)}.$$

Equivalently, we state that alternative a dominates alternative b with respect to the class $V(I)$. Note that the resulting relation is no longer complete: alternatives exist for which neither one nor the other direction of dominance is given. For these pairs of alternatives, we have "incomparability", which has to be distinguished from indifference $\sim_{V(I)}$. The latter case would by definition only occur if there was indifference between a and b for all v in $V(I)$.

If the concept of incomplete information is to be applied, there are two problems that must be solved. First, the class of feasible value functions needs to be determined. Second, it is necessary to examine for the relevant pairs of alternatives whether one alternative dominates the other with respect to the class of value functions. At this point, we should explain the procedures used to determine the class and to check for dominance in general terms as well as by an example: however, we would like to postpone doing so until later. For value functions with one objective, the procedures are not particularly interesting and only of minor practical relevance. This will be different when we apply the concept of incomplete information to decision making with multiple objectives, decision making under risk and group decisions. We will explain these procedures in detail in the respective chapters.

Questions and exercises

5.1
What is the difference between ordinal and measurable value functions?

5.2
An entrepreneur has made the following statements concerning his value function: the value of the difference between -€10M and -€7M is the same as the values of the transitions from -€7M to -€5M, from -€5M to -€3M, from -€3M to €0M, from €0M to €1M, from €1M to €5M, and from €5M to €10M.

(a) Draw the value function normalized on the interval [0, 1].
(b) Which method is used?
(c) Check for consistency.

5.3
You are looking for a new flat. One of the important considerations is the size of the living space. You have offers of flats with living spaces between 20m^2 and 200m^2.

(a) Assume you always prefer the flat with the larger living space. You do not care about the disadvantages of bigger flats – more cleaning, higher rent, higher heating costs – as your grandma will pay these expenses. Determine the value function by using the bisection method.
(b) Assume now that your grandma does not handle all the related expenses. What is the resulting problem and how can you solve it?

5.4
Think again about the entrepreneur from Exercise 5.2. The relevant range of profit was between -€10M and €10M. The entrepreneur makes the following statements: the value of the difference between -€10M and -€7M is the same as values of the transitions from -€7M to -€5M, from -€5M to -€3M, from -€3M to €0M, from €0M to €1M and from €1M to €5M. However, the transition from €5M to the maximum profit of €10M creates a smaller value than the other transitions. Can you still generate the value function for the profits from -€10M to €10M if the entrepreneur makes one of the following statements?

(a) The transition from €1M to €5M generates the same value as the transition from €5M to €12M.
(b) The transition from €1M to €5M generates the same value as the transition from €5M to €12M. Within the range between €5M and €12M, the transition from €5M to €10M generates the same value as the transition from €10M to €12M.
(c) The transition from €5M to €10M generates half as much value as the transition from €1 to €5M.
(d) The transition from €5M to €10M generates the same value as the transition from -€7M to -€6M.

5.5
(a) Construct a value function with the target variable "distance between my flat and the city" (where you work and shop, but where pollution and noise are both bothersome).
(b) In the case that your value function is not monotonous: can you substitute the attribute "distance" by several different attributes for which you have a monotonous value function?

5.6
You think about giving money to your friend as a wedding present. Therefore, you determine your friend's value function for money in the range [€0, €1,000] by a hypothetical questioning (i.e. you anticipate the answers he might have given to your questions when applying two different methods).

The result of method I is $v_I(x) = x/500 - x^2/1{,}000{,}000$; method II yields $v_{II}(x) = 1{,}582 - 1{,}582\,e^{-0{,}001 \cdot x}$.

What are you going to do now?

Chapter 6: Decision making under certainty and with multiple objectives: multiattribute value functions

6.0 Summary

1. In many situations, multiple objectives are relevant when making a decision. They require the evaluation of alternatives by means of a multiattribute value function.
2. The simplest and most important multiattribute value function is the additive one. Using this value function, the (total) value of an alternative is computed as a weighted sum of (individual) values per attribute.
3. The additive model can only be employed if certain independence conditions between the considered attributes are fulfilled.
4. The weights of the attributes can reasonably be determined using the *trade-off* method or the *swing* method. In the case of the trade-off method, objective weights are derived from the trade-off relationship between two attributes. In the case of the swing method, objective weights are computed from point valuations of different alternatives.
5. The widely used *direct-ratio* method is fraught with problems.
6. Attribute weights can only be reasonably interpreted with respect to the ranges of the attribute levels. There is no "per se" importance of particular attributes.
7. If the decision maker is incapable of specifying exact weights, the optimal alternative might also be determined on the basis of imperfect information (e.g. the specification of an interval) or at least dominant alternatives can be eliminated.
8. Sensitivity analyses can provide important information on acceptable changes of the objective weights without changing the optimal alternative.
9. Individuals make systematic mistakes when trying to determine their attribute weights. In particular, they do not sufficiently take into account the attribute ranges (see 6. above) and overestimate the weight of an objective if it is split into multiple subordinate objectives.

6.1 Value functions for multiple attributes

Decision making with multiple conflicting objectives is a key problem in many areas of application. For that reason, the literature on this topic is particularly multifaceted and extensive (for a recent survey, see Wallenius et al. 2008). We will focus on the multiattribute valuation concept in this chapter. Building on the ideas in Chapter 5, in which we analyzed value functions for only one objective, we will now consider value functions for multiple attributes. A multiattribute value func-

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_6, © Springer-Verlag Berlin Heidelberg 2010

tion assigns a particular value to every alternative, depending on the respective levels of the individual attributes. We assume certainty concerning the levels.

Let us return to the example from Chapter 5 and consider the recent graduate who faces several employment opportunities. They no longer differ only in their respective annual starting salary but now also in their average weekly working time: occupation in a management consultancy, as a teaching assistant at a university, or as a sailing instructor (Table 6-1).

Table 6-1: Three employment opportunities with two attributes

Alternative	Salary	Working hours
(*a*) Consultancy	€80,000	60 hours
(*b*) University	€50,000	40 hours
(*c*) Sailing instructor	€30,000	20 hours

Using a *multiattribute value function v*, we want to try to model his preferences according to his multiple objectives (i.e. salary and working hours) in order to facilitate the complex decision. (Having acquired the relevant knowledge, he will be able to do so himself.) Similarly to the value functions for a single attribute, the value function v shall express the strength of preference towards the different alternatives. Among two alternatives, the one with the higher (preference) value is chosen. We are hence looking for a function v with

$$a \succ b \Leftrightarrow v(a) > v(b).$$

But how do we obtain such a function? First off, there needs to be a guarantee that such a function exists in the first place. This does not mean that the decision maker needs to know it; the "existence" of such a function rather means that the decision maker is capable of making statements that allow the construction of the function.

Let us assume that the value function v exists. What could it look like? Obviously, a particularly simple form, typical of the additive model, is desirable. We will thus restrict our attention to this model, especially as it has the highest relevance in practice. In Section 6.3, we will discuss what conditions are necessary for the existence of an additive multiattribute value function.

6.2 The additive model

The alternative $a \in A$ is characterized by the vector $a = (a_1, \dots, a_m)$. The a_r indicate the levels of the attribute X_r for the alternative a. The decision maker has a value function $v_r(x_r)$ (one-dimensional attribute value function, individual value function) over every attribute X_r. The individual value functions v_r are normalized on the interval [0, 1]. This means that all relevant levels lie between x_r^- (worst level) and x_r^+ (best level) and that

$$v_r(x_r^-) = 0 \quad \text{and} \quad v_r(x_r^+) = 1 \tag{6.1}$$

hold.

The boundaries x_r^- and x_r^+ have to enclose the attribute levels of all alternatives that are being evaluated but they can be larger than necessary (e.g. to avoid re-normalizations in the case of the occurrence of new alternatives). However, the interval should generally be as narrow as possible as this leads to more accuracy in the subsequent evaluations. If, for instance, your job offers yield salaries of €55,000 and €68,000, it is possible to determine the value function over this interval or over a slightly larger one, say from €50,000 to €80,000; however, it would not be sensible to use an interval from €0 to €1,000,000.

The additive model determines the value of an alternative a as

$$v(a) = \sum_{r=1}^{m} w_r v_r(a_r) \tag{6.2}$$

where $w_r > 0$ and

$$\sum_{r=1}^{m} w_r = 1. \tag{6.3}$$

As stated above, a_r indicates the level of the attribute X_r for the alternative a, and $v_r(a_r)$ indicates the respective value of the attribute value function v_r. The w_r are objective weights or attribute weights. Later on, we will explain that this term is misleading because attributes actually do not have any weights per se. However, we will use this term because it is both convenient and widely used, although it would be more accurate to speak of scaling constants. The weights relate the valuations of the different attributes to each other. Essentially, only the relative size of the weights is important, and not their absolute size. The normalization that is postulated by condition (6.3) has the single purpose of choosing one specific combination out of the set of many equivalent weight combinations allowing us to speak of *unique weights*. In multiattribute value theory, it is common to normalize by fixing the sum of weights at 1 (and this is actually in line with our standard comprehension of "weighting"). However, we could as well postulate that w_1 always has to be 1 (a normalization that you will observe later on in Chapter 11, when we discuss intertemporal decision making and interpret it as a special case of a multiattribute decision problem).

If you consider just two attributes X and Y, the additive model can be illustrated graphically. Figure 6-1 shows how the total value is aggregated from the attribute value functions v_X and v_Y and the weighting factors w_X and $w_Y = 1 - w_X$. x^- indicates the worst, x^+ the best level of the attribute X; the same applies to Y. We obtain the left (total) value function if attribute X is weighted heavily and the right one if a lot of weight is assigned to attribute Y. The meaning of the weighting factors can easily be seen: the weighting factor w_X indicates the increase in value when attribute X is changed from its lowest level (individual value = 0) to the highest level (individual value = 1), leaving all other attributes equal. In general,

the value of an alternative with the maximum level in one attribute w_r and the minimum level in all other attributes is equal to the weight of this attribute:

$$\begin{aligned} & v(x_1^-, x_2^-, \ldots, x_r^+, x_{r+1}^-, \ldots, x_m^-) \\ &= w_1 v_1(x_1^-) + w_2 v_2(x_2^-) + \ldots + w_r v_r(x_r^+) + w_{r+1} v_{r+1}(x_{r+1}^-) + \ldots + w_m v_m(x_m^-) \\ &= w_1 \cdot 0 + w_2 \cdot 0 + \ldots + w_r \cdot 1 + w_{r+1} \cdot 0 + \ldots + w_m \cdot 0 \\ &= w_r. \end{aligned}$$

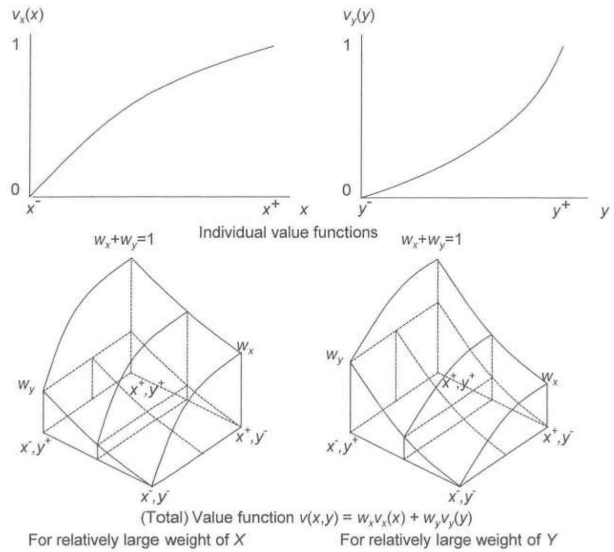


Figure 6-1: Graphical illustration of the additive model for two attributes

Let us assume we had identified measurable attribute value functions v_1 and v_2 for our job example (using the techniques described in Chapter 5). Let us further assume that the value functions in this case are monotone. This implies that the decision maker prefers more money to less money and less work to more work at all times (this is a quite plausible assumption but need not invariably be the case).

The value functions in Table 6-2 were determined over the attribute intervals €30,000 to €80,000 annual salary and 20 to 60 weekly working hours, respectively.

Table 6-2: Three job offers with two attributes and their respective values

Alternative	Salary x_1	Value of salary $v_1(x_1)$	Working hours x_2	Value of working hours $v_2(x_2)$
(*a*) Consultancy	€80,000	1.0	60 hours	0.0
(*b*) University	€50,000	0.6	40 hours	0.5
(*c*) Sailing instructor	€30,000	0.0	20 hours	1.0

With the help of the value function v, the total value can now be derived from the individual values. Let us assume that we know both objective weights w_1 and w_2. This allows us to evaluate the three alternatives. From Table 6-3, we see that for $w_1 = 0.6$ and $w_2 = 0.4$, the job as a consultant yields the highest total value.

Table 6-3: Evaluation of the three job offers using attribute weights of 0.6 for the salary and 0.4 for working hours

Alternative	Value of salary $v_1(x_1)$	Weighted value of salary $w_1v_1(x_1)$	Value of working hours $v_2(x_2)$	Weighted value of working hours $w_2v_2(x_2)$	Total value $w_1v_1(x_1) + w_2v_2(x_2)$
(*a*) Consultancy	1.0	0.60	0.0	0.00	0.60
(*b*) University	0.6	0.36	0.5	0.20	0.56
(*c*) Sailing instr.	0.0	0.00	1.0	0.40	0.40

6.3 Requirements for the applicability of the additive model

Due to the simplicity and elegance of the additive value function, the model is (often hastily) used in a variety of applications. For example, it is known by the terms "scoring model", "point valuation method", or "cost utility analysis". In these methods, scores between 0 and 10 or 0 and 100 are attached to the attribute levels of each alternative. The importance of an attribute is reflected by a weight expressed as a percentage. The percentages add up to 100% for all attributes. The percentages are used to weight the scores, summing up to the total value of an alternative. Examples can be found in a variety of domains, e.g. in product testing, methods of analytical job performance evaluation, the evaluation of alternative technical systems, performance evaluations in sports, or even at school.

In order to be able to rationally justify the use of an additive value model, certain conditions concerning the independence of the attribute evaluation have to be fulfilled. A measurable value function

$$v(x_1, x_2, \ldots, x_m) = w_1v_1(x_1) + w_2v_2(x_2) + \ldots + w_mv_m(x_m)$$

obviously expresses the concept that a certain increase in one attribute causes a change in total value completely independent of the level of all other attributes. For example:

- In a decathlon, the improvement from 11.5 to 11.0 seconds in the 100-meter sprint yields an additional score that is independent of the performance in the long jump or shot put event.
- In building projects, short construction time, a high quality standard, and low construction costs are desirable. The reduction of the construction time from 10 to 9 months generates the same value for you, no matter whether costs and quality are high or low.

We have already discussed this property in Chapter 3, where we called it preferential independence. Let us now define it formally:

Definition 6.1 (Simple preferential independence)
Let

$$a = (a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_m)$$
$$b = (a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_m)$$

be two alternatives that only differ in the i-th attribute, and

$$a' = (a_1', \ldots, a_{i-1}', a_i, a_{i+1}', \ldots, a_m')$$
$$b' = (a_1', \ldots, a_{i-1}', b_i, a_{i+1}', \ldots, a_m')$$

be two other alternatives that also only differ in the i-th attribute, but for this feature the same levels as a and b. The attribute X_i is then called *(simply) preferentially independent* of the remaining attributes if it holds for any a, b, a', b' defined as above that:

$$a \succ b \Leftrightarrow a' \succ b' \quad \text{(analogous for } \sim, \prec\text{)}.$$

To clarify: you want to buy a new car and choose among various models. For you, one of the most important attributes is the color and you are only interested in a black or a white car. For every model, both colors are available. If you prefer black on *any* car model, the attribute "color" is simply preferentially independent of all other attributes. However, if you preferred an Opel in black and a VW in white, the attribute color would not be independent of the other attributes.

Extending the definition of preferential independence to the case of multiple attributes does not pose a problem. A subset of the attribute set is called "preferentially independent of the remaining attributes" if the preference of the decision maker with respect to the attribute levels on this subset is not influenced by the levels of the other attributes (as long as they are the same for all alternatives). We refrain here from presenting a formal definition of this property (it would be a little technical and rather confusing) and explain the extended concept by means of an example:

A professor is looking for an undergraduate research assistant and a number of students have applied for the job. After a first screening, two of them remain on the short list. The relevant objectives are grade point average (GPA), length of study, knowledge of foreign languages, computer skills, and maximum workload in hours per week. An assistant is asked to support the preparation of the final decision by summarizing the relevant attribute levels for both candidates. He constrains his summary to the two attributes of grade point average (3.5 vs 3.0) and maximum workload (3 hrs/w vs. 8 hrs/w) and justifies his approach with the argument that "the candidates do not differ on the other attributes anyway". However, in doing so he has implicitly assumed that the attributes GPA and maximum workload are preferentially independent of the other attributes and that the explicit levels of the other attributes do not even need to be mentioned. Whether this is a sensible assumption is very doubtful. It may well be that the difference in maximum workload is much more important (relative to the GPA) if both candidates have extraordinary computer skills, while it is less relevant if both candidates are rather weak in this respect.

We want to stress once more the fundamental difference between preferential independence and statistical independence. Preferential independence has nothing to do with the empirical observation that students with better computer skills have also better GPAs and less time available, for instance (the authors have not checked whether this is indeed an empirical fact; it is only an example), it is instead concerned with the feature that the decision maker has a preference with respect to different GPA/workload combinations that is not influenced by other attributes like computer skills.

A particularly comfortable decision situation is given if the above defined condition of preferential independence is given for any subset of the attribute set; we will call such a scenario "mutual preferential independence".

Definition 6.2 (Mutual preferential independence)
The attributes $X_1, \ldots, X_m$ are *mutually preferentially independent* if every subset of these attributes is preferentially independent of its complementary set of attributes (R. L. Keeney and Raiffa 1976, p. 111).

Let us look at an example: if you want to buy a new car and consider the four criteria (attributes) "engine power", "trunk size", "consumption", and "price" for your evaluation, the condition implies that the subset {engine power, trunk size, consumption} should be preferentially independent of the complementary set {price}. At the same time, the subset {engine power, price} has to be preferentially independent of the complementary set {trunk size, consumption} and so on (for each possible subset of attributes).

If mutual preferential independence is satisfied, the preference can be represented by an additive multiattribute value function. However, mutual preferential independence is only sufficient for non-measurable value functions. These are value functions that are able to arrange the alternatives in the right order, but whose values cannot be used to measure value differences ("strength of preference"). Since we would like to obtain measurable value functions (which means that we want to be able to interpret value differences), preferences have to meet even stricter requirements.

Definition 6.3 (Difference independence)
Let

$$a = (a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_m)$$
$$b = (a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_m)$$

be two alternatives that only differ in the i-th attribute, and

$$a' = (a'_1, \ldots, a'_{i-1}, a_i, a'_{i+1}, \ldots, a'_m)$$
$$b' = (a'_1, \ldots, a'_{i-1}, b_i, a'_{i+1}, \ldots, a'_m)$$

be two more alternatives that also only differ in the i-th attribute, but for this feature the same levels as a and b. The attribute X_i is then called *difference independent* of the remaining attributes if for any a, b, a', b' defined as above it holds that

$$(a \to b) \sim (a' \to b').$$

Let us look at an example: the attribute "maximum speed" is difference independent of the attributes "price" and "consumption", if the additional value that you attach to a particular increase in maximum speed is independent of the vehicle being a €50,000 car with an average consumption of 20 liters/100km or a €20,000 car with an average fuel consumption of 10 liters/100km.

An additive measurable value function requires that additive difference independence holds for every attribute; it is easy to verify that this also implies mutual preferential independence.

Hence, our earlier statement that "preferential independence" is the necessary condition for the additive model was very sloppy. More exactly, we have to require mutual preferential independence and, if we ask for a measurable value function (as we usually do), we also need difference independence.

If it is impossible for a decision maker to define the attribute value function of an attribute without knowing the level of another attribute, preferential independence is obviously not given. Let us assume that, in a decision between two jobs, both the annual salary and the annual number of days off are important besides other attributes. The decision maker now wants to construct his attribute value function for the annual number of days off and makes use of the bisection method. He hence has to determine the number of days off that is right in the middle of 20 and 40 vacation days from a value perspective. He should now ask himself: is it possible to determine this without knowing my salary? If I have a very low salary (€35,000 gross), I have to spend a large part of my net income on daily life goods so there will not remain much money for holidays. Therefore, holiday time exceeding three weeks is of very little value. Even though I can read or take long walks, I will probably soon miss the stimulating atmosphere at work. The midpoint level from a value perspective might be about 24 days. However, if my salary is close to the upper limit of the considered alternatives (€60,000 gross), additional holiday time is of much greater value (luxurious long-distance journeys). The midpoint level from a value perspective might then be about 29 days.

The difference independence condition will tend to be fulfilled (approximately) more easily if the ranges between the upper and lower limits of the respective attributes are small. If the minimum salary is €50,000 and the maximum salary is €60,000, it can be expected that the value function over the days off is less dependent on the salary in comparison to a range from €30,000 to €60,000.

Figure 6-2 shows two *non-additive* value functions over two attributes X and Y. In the left example, the increase in value when switching from x^- to x^+ is quite modest at a low level of Y. The higher Y becomes, the more value appreciation is produced by the increase in the level of X; there is a complementary relationship between the attributes. Consider a medical decision and let X denominate the life duration and Y the quality of life. If we (plausibly) assume that the decision maker values an additional year of life more if he enjoys a higher quality of life, we have complementarity. The reverse case is displayed in the right part of Figure 6-2: Here, we observe a substitute relationship between the attributes. The higher the level of one attribute, the less additional value is produced by an improvement of the other attribute.

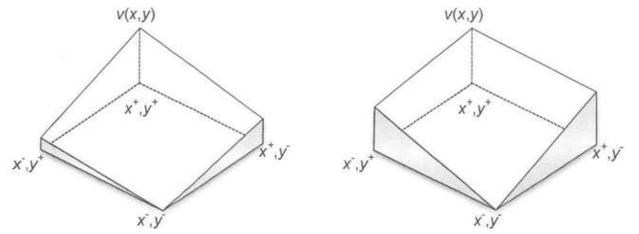


Figure 6-2: Non-additive value models with two attributes

If the conditions for the additive model are not fulfilled, you should – as already mentioned in Chapter 3 – try to "produce" independence via a different, better formulation of the objectives. For example, you could split the attribute "number of days off" into three sub-attributes:

- Number of days off for long-distance journeys,
- Number of days off for trips within Europe,
- Number of days off spent at home.

For each of these attributes, you could generate a value function. In addition, you would have to decide how you plan to assign your days off to the three categories, depending on your annual salary.

Another example: a manager who is in charge of a profit center tries to plan the development of his unit over the next five years. For him, not only the expected total profit over this time span is important but also the development of the profits

over time. He prefers a continuous upward trend to a volatile up-and-down movement that comes, in the worst case, with a sharp decline in the final year. If the profit in each of the five years was introduced as a separate attribute, there would be no difference independence between them. However, it might be possible to model the manager's preferences through the two attributes, say "total profit over five years" and "average annual profit growth", which fulfill the independence condition.

Since an additive value model strongly simplifies the decision calculus, it is worthwhile to carefully think about possibilities to redefine attributes in order to eliminate existing dependencies. Von Winterfeldt and Edwards, who are not only scientists but also experienced decision analysis practitioners, are convinced that this should be possible in virtually every case of application ({{324 von Winterfeldt,Detlof. 1986/y;}},1986, p. 309).

6.4 Determination of the weights

6.4.1 The attribute value functions in the example "Choosing a job"

From now on, we will always assume that preferences can be properly modeled by a measurable, additive value function. There are a number of possible ways to determine the attribute weights w_r. We will outline three of these approaches and illustrate them by means of examples. In order to do so, we extend the above-mentioned job choice problem and introduce a third objective, which we will call "career perspectives". The respective attribute can assume the levels "excellent", "good", and "bad". We do not want to go into further detail regarding the measurement and interpretation of these terms. However, it is important to note that the decision maker himself has to have a clear understanding of their meaning; otherwise he will not be able to determine a proper attribute value function and an appropriate weighting of this attribute.

Table 6-4 displays the extended alternatives. The value function of the decision maker for the starting salary, v_1, is assumed to have a shape as presented in Figure 6-3. The elicited points were approximated by the continuous function

$$v_1(x_1) = 1.225\cdot(1-e^{-1.695(x_1-30{,}000)/(80{,}000-30{,}000)}).$$

Table 6-4: The three available jobs with three attributes

Alternative	Salary	Working hours	Career perspectives
(*a*) Consultancy	€80,000	60 hours	good
(*b*) University	€50,000	40 hours	excellent
(*c*) Sailing instructor	€30,000	20 hours	bad

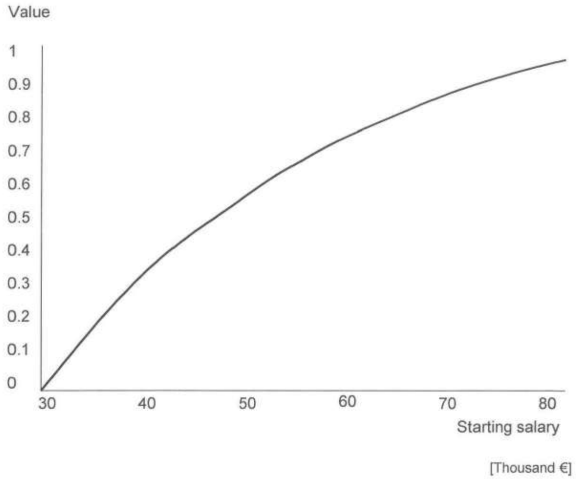


Figure 6-3: Value function for an annual salary in a range between €30,000 and €80,000

For the attribute "working hours", the decision maker is assumed to have a linear value function in a range between 20 and 60 hours; it thus holds that $v_2(x_2) = (60 - x_2) / 40$. For the attribute "career perspectives", the attribute levels are evaluated according to the last column of Table 6-5.

Table 6-5: Individual values of the alternatives in the respective attributes

Alternative	Salary	Working hours	Career perspectives
(*a*) Consultancy	1	0	0.7
(*b*) University	0.6	0.5	1
(*c*) Sailing instructor	0	1	0

6.4.2 Determination of the weights by use of the trade-off method

A *trade-off* is essentially an exchange relationship. Determining the weights according to the trade-off method means that you ask for the exchange relationship of two objective variables for which the decision maker is indifferent. The value functions have to be known. You proceed as follows: you need pairs of alternatives that differ only in two attributes and are considered equally attractive by the decision maker. An indifference statement, for instance, between the alternatives

$$f = (f_1, \ldots, f_{i-1}, f_i, f_{i+1}, \ldots, f_{j-1}, f_j, f_{j+1}, \ldots, f_m)$$

and

$$g = (f_1, \ldots, f_{i-1}, g_i, f_{i+1}, \ldots, f_{j-1}, g_j, f_{j+1}, \ldots, f_m)$$

can be used to conclude how heavily the decision maker weights the attributes X_i and X_j. Due to the additivity, the equation $v(f) = v(g)$ can be reduced to

$$w_i v_i(f_i) + w_j v_j(f_j) = w_i v_i(g_i) + w_j v_j(g_j).$$

A set of $m - 1$ of such equations in combination with the normalization condition $\Sigma w_r = 1$ yields a system of m equations in m variables. This system has a unique solution if there are no redundancies between the equations. Redundancies occur if you can derive an equation of the system from two or more other equations within the system.

Let us begin by comparing the attributes "annual salary" and "working hours". Since it is often very difficult for a decision maker to immediately quantify precisely an exchange relationship, you would prefer to draw nearer step-by-step, starting with the extreme values. At first, the decision maker is asked if he prefers an alternative f with the characteristics

$$f = (€80{,}000, 60 \text{ hours}, *)$$

which is the combination of highest possible salary and worst outcome in terms of working hours, to an alternative g with

$$g = (€30{,}000, 20 \text{ hours}, *)$$

which is the combination with the worst possible salary and the best working hours. The attribute level for career perspectives is omitted here, because it is irrelevant for the preference determination as long as the level is the same for both alternatives, due to the assumed preferential independence; that is why an asterisk has been put in the third place. It is sufficient for the decision maker to focus solely on the first two attributes when determining the *trade-off*.

If the decision maker states, for instance, that he prefers alternative f to g, then the high salary of €80,000 can be lowered to €60,000, whereupon the decision maker has to decide between f' and g with

$$f' = (€60{,}000, 60 \text{ hours}, *)$$
$$g = (€30{,}000, 20 \text{ hours}, *)$$

If he still prefers f', the salary of €60,000 can be reduced further. If his preference reverses to g, the salary of €60,000 will be increased again. This process is repeated a couple of times until a sum of €x with $f'' = (€x, 60 \text{ hours}, *)$ is found that makes f'' and g equally attractive. Let us assume €55,000 is the amount we are looking for. Hence it holds

$$(€55{,}000, 60 \text{ hours}, *) \sim (€30{,}000, 20 \text{ hours}, *)$$

meaning that the decision maker is willing to trade an additional salary of €25,000 (on top of the €30,000) against 40 hours of weekly working time (on top of the 20 hours); this is his trade-off.

It holds that

$$w_1 v_1(€55{,}000) + w_2 v_2(60\text{ hours}) + w_3 v_3(*) =$$
$$w_1 v_1(€30{,}000) + w_2 v_2(20\text{ hours}) + w_3 v_3(*).$$

Since * is equal in both cases, the term can be canceled out on both sides. Solving for w_1 yields

$$w_1 = \frac{v_2(20) - v_2(60)}{v_1(55{,}000) - v_1(30{,}000)} \cdot w_2 = \frac{1}{0.7} \cdot w_2 = 1.429 \cdot w_2 . \tag{6.4}$$

The approximate value for the salary $v_1(55{,}000) = 0.7$ can be taken from Figure 6-3. The exact value is derived by inserting the argument into the given exponential function.

We now compare the attributes "career perspectives" and "annual salary". Let us assume the decision maker arrives (after an appropriate convergence procedure) at the following equally attractive alternatives:

$$(€70{,}000, *, \text{bad}) \sim (€30{,}000, *, \text{excellent})$$

This would lead us to the following equation:

$$w_1 = \frac{v_3(\text{excellent}) - v_3(\text{bad})}{v_1(70{,}000) - v_1(30{,}000)} \cdot w_3 = \frac{1}{0.91} \cdot w_3 = 1.099 \cdot w_3 . \tag{6.5}$$

Adding the normalization condition $\Sigma w_r = 1$, we can easily derive the respective w_r from the equations. We find $w_1 = 0.38$, $w_2 = 0.27$, and $w_3 = 0.35$.

The decision maker can use these weighting factors to obtain a ranking of his preferred alternatives:

$$v(a) = 0.38 \cdot 1 + 0.27 \cdot 0 + 0.35 \cdot 0.7 = 0.63$$
$$v(b) = 0.38 \cdot 0.6 + 0.27 \cdot 0.5 + 0.35 \cdot 1 = 0.71$$
$$v(c) = 0.38 \cdot 0 + 0.27 \cdot 1 + 0.35 \cdot 0 = 0.27.$$

As we can see, the occupation as a teaching assistant at the university is the optimal alternative, whereas the job as a sailing instructor ends up far behind the two other alternatives.

One potential problem with the determination of trade-offs can be caused by attributes that assume only a few possible levels. There is no continuous value function. This can be problematic for the reason that the level of a particular attribute can only be adjusted discontinuously. For example, the attribute "location" could play an important role in choosing a job. The head office of the consultancy is in Frankfurt, the university is located in Paderborn, and the sailing academy in Kiel. The decision maker has assigned a value of 1 to his preferred location (e.g. Kiel), a value of 0 to his least preferred location (e.g. Paderborn), and

an intermediate value to the third location (Frankfurt). When comparing the location with a continuous attribute like the starting salary, it is no problem to find equally attractive alternatives because it is possible to vary the level of the continuous attribute. If, for instance, the decision maker states (Paderborn, €80,000) $\succ$ (Kiel, €30,000), then you can increase the salary that is associated with the location Kiel until the decision maker is indifferent between the alternatives. However, you should avoid a comparison of two discrete attributes because it would only be by mere chance that you would find two alternatives with the same attractiveness.

Moreover, it may happen that there is such a large preference jump between two adjacent levels of the discrete attribute that it cannot be compensated for by any adjustment of the continuous attribute within its predefined range. In our case, e.g., the number of annual holidays (as another attribute) may play a role as well. The number of holidays has a very narrow range of variation; let us say 20 to 24 days. Beginning with a combination (Kiel, 20 days), it might not be possible to compensate the transition to a less preferred location by increasing the number of holidays. Hence, there are no two combinations within the given range of alternatives that are equally attractive. However, since we are hardly restricted in our choice of picking attribute pairs to construct the trade-offs (only keeping the problem of redundancies in mind), we should not encounter a problem here.

In our sample calculations, trade-offs were always determined on the basis of alternatives that featured only minimal or maximal levels of an attribute. It is problematic that the decision maker is thereby confronted with rather unrealistic alternatives. He has to be imaginative and will have a harder time making reliable indifference statements.

You should therefore try to use realistic comparisons when identifying trade-offs. In a comparison between two different combinations of salary and working hours, the decision maker may well be able to make the following statement:

$$(€80{,}000, 60\text{ hours}, *) \sim (€65{,}000, 52\text{ hours}, *)$$

The left alternative is equivalent to the alternative "consultancy" in both considered attributes. It should be fairly easy for the decision maker to imagine this combination. From the indifference statement we can derive the following relation:

$$w_1 = \frac{v_2(52) - v_2(60)}{v_1(80{,}000) - v_1(65{,}000)} \cdot w_2 = \frac{0.2}{0.15} \cdot w_2 = 1.333 \cdot w_2 .$$

You can see that this weight relation does not exactly match the relation ($w_1 = 1.429\, w_2$) that was determined before. However, if the additive model was valid (as we presume) and if the attribute value functions were properly elicited beforehand, we should have obtained the same ratio in both cases. Differences of such kind are inevitable due to people's limited cognitive abilities. For that reason, it is sensible not to limit ourselves to the determination of $m - 1$ trade-offs to generate a consistent system of equations but to deliberately create more than the minimal required number of trade-off measurements in order to check for consistency of the statements. In Section 6.5, we will investigate ways and means of dealing with the unavoidable fuzziness of preference measurements.

6.4.3 Determination of the weights by use of the swing method

Unlike the trade-off method, the swing method does not require that the attribute value functions are known. The decision maker imagines that he has at hand the worst defined alternative

$$a^- = (x_1^-, x_2^-, \ldots, x_m^-).$$

Assume that he has the choice of increasing one of the attributes to its highest level, leaving all other attributes at their lowest levels. He ranks the attributes according to his preference for increasing their level to the maximum. Let us call these artificial alternatives

$$b^r = (x_1^-, \ldots, x_{r-1}^-, x_r^+, x_{r+1}^-, \ldots, x_m^-)$$

with $r = 1, \ldots, m$ and remind ourselves that $v(b^r) = w_r$.

Through the ranking of the b^r, we have already structured the weights in descending order; it now remains to quantify them. Let us assign a value of zero to a^- and an arbitrary value of 100 points to the most preferred b^r. Subsequently, the decision maker assigns points to the remaining b^r in such a way that the value differences between them are reflected. The final step is the normalization of the weights to 1. Let t_i be the number of points, it then holds that

$$w_r = \frac{t_r}{\sum_{i=1}^{m} t_i} \qquad (6.6)$$

In summary, the procedure works as follows:

1. Specification of a ranking of the artificial alternatives b^r,
2. Base allocation of points: 0 for a^-, 100 for the best b^r,
3. Evaluation of the remaining b^r in such a way that the value differences between them are reflected,
4. Determination of the attribute weights by normalization of the evaluations.

The approach can be clarified by means of our example. At first, the decision maker considers the extreme alternative

$$a = (€30{,}000, 60 \text{ hours, bad}).$$

He is then asked for which attribute a change from the lowest to the highest level would be most attractive for him. He hence has to decide which of the alternatives

$$b^1 = (€80{,}000, 60 \text{ hours, bad})$$
$$b^2 = (€30{,}000, 20 \text{ hours, bad})$$
$$b^3 = (€30{,}000, 60 \text{ hours, excellent})$$

he prefers. Let us assume he chooses b^1 as his preferred alternative, followed by b^3 and finally b^2.

Let us further assume that the decision maker arrives at an evaluation as depicted in Table 6-6.

Table 6-6: Evaluation of the three artificial alternatives b^r

Rank	Alternative b^r	Points
1	b^1	100
2	b^3	70
3	b^2	60

We then derive the weights:

$$w_1 = 100 / (100 + 70 + 60) = 0.44$$
$$w_2 = 60 / (100 + 70 + 60) = 0.26$$
$$w_3 = 70 / (100 + 70 + 60) = 0.30.$$

The calculation of the weights from the preference statements by the swing method is fairly easy. However, it should be noted that – in the same spirit as in the comparison of the value function elicitation methods in Section 5.2.5 –the allegedly simpler swing method actually requires much more cognitive work on part of the decision maker than does the trade-off method. While the trade-off method asks for a large number of very simple preference statements (i.e., "I like this combination better than that one"), the "point assignment by preference strength" (step 3) of the swing method is a very complex valuation problem that additionally leaves a large margin of interpretation (what exactly does "assigning points by preference strength" even mean?). Therefore, the major advantage of the swing method should not be seen in the property that the weights are generated in such an easy way (with respect to quality and reliability of the elicited weights this should actually be seen as detrimental) but rather in the fact that the detailed shapes of the attribute value functions are not needed in the process (as only extreme levels are compared).

6.4.4 Determination of the weights by use of the direct-ratio method

In practice, the direct-ratio method is widely used (even though it is often referred to by another name). Strictly speaking, it should not even be presented in a textbook because it is quite unreliable. However, we must outline this method for two reasons: first, it is very widely used in practice and second, we want to clarify its logical defect.

When using this method, you first have to rank the attributes according to their "importance". This does not cause too much trouble for most people, although a question concerning the importance of an attribute per se is pointless. The objective variable itself cannot be important, only the difference between its levels. For example, it is pointless to say that salary is more important than holidays. The following statement would be more sensible: an increase of annual salary from €50,000 to €53,000 is more important to me than an increase of annual holiday time from 25 to 30 days. We will come back to this issue in Section 6.6.

Let us assume that you still have the feeling that the annual salary is more important to you than your career perspectives and that your career perspectives are more important than the working hours. You now compare two attributes at a time; let us start with the less important ones. The question is: "how much more important to you are your career perspectives than the working hours?" Your answer might be "just a bit". This answer is of little use and you are therefore asked to attach a concrete figure to it. The wording of the question is now as follows: "If the level of importance that is attached to the attribute "working hours" equals 1, how important is the attribute "career perspectives"?" Your answer might be "1.2". You proceed accordingly when comparing the attributes "annual salary" and "working hours". "If the level of importance that is attached to the attribute "working hours" equals 1, how important is the attribute "annual salary"?" Your answer: 2.

From these statements you can derive objective weights. It holds that

$$w_1 / w_2 = 2$$

and

$$w_3 / w_2 = 1.2.$$

It follows for the weights that

$$w_1 = 2/(1+1.2+2) = 0.48$$
$$w_2 = 1/(1+1.2+2) = 0.24$$
$$w_3 = 1{,}2/(1+1.2+2) = 0.29.$$

Obviously, it is again advisable to check for consistency by comparing the importance of the attributes salary and career perspectives. In this comparison, the decision maker would be consistent if he quoted a ratio of $w_1/w_3 = 2/1.2 = 1.7$.

6.4.5 Application of multiple methods and alternative procedures

As has been repeatedly emphasized, it is advisable to attempt to validate the statements that lead to the determination of weights. This could happen within a single method, but multiple methods could be used as well. For example, you could identify the weights using the trade-off method and test these weights with the swing method, or vice versa.

An alternative method that has some appeal for practical applications of multiattribute decision making was developed by Hammond et al. (1998). It is called the *even-swap* method. The procedure differs fundamentally from the methods discussed so far as it does not determine attribute value functions and attribute weights in isolation; however, the even-swap method builds on the same conceptual ideas as the trade-off method, in particular on the assumption that mutual preferential independence allows the comparison of changes on some attribute levels without concern for the levels of the other attributes. More explicitly, the even-swap method at each step derives changes in exactly two attributes (one improving and one worsening) that leave the decision maker's preference unchanged. In contrast to the trade-off method, however, these even swaps are not used to derive attribute weights but to transform a given alternative into an equally attractive alternative. This is repeated again and again with other even swaps until obvious dominance relations can be spotted (see the sample case at the end of Chapter 6). In this process, the (hypothetical) alternatives that are successively generated by the even swaps are only considered for easy comparison. After the optimal alternative in the set of modified alternatives has been identified, we carry over the preference ordering to the original alternatives.

The even-swap method has the advantage that it does not need terms like attribute value functions and attribute weights even though it is based on exactly these concepts. Furthermore, the method is easy to explain and to understand; in practical applications, this can be an important advantage. And, if supported by appropriate software, even the search for dominance relations can be arranged in a very efficient way (Mustajoki and Hämäläinen 2005). For a deeper analysis of complex problems, the classical approach of determining value functions and attribute weights can hardly be circumvented.

6.5 Incomplete information about the weights

6.5.1 Handling of inconsistent or incomplete information

Usually, the generation of redundant information will lead to inconclusive weights due to inconsistencies in the statements of the decision maker. He should become aware of these inconsistencies and reassess and revise his statements until they are consistent. However, it is also possible that the decision maker is not willing or capable of correcting the inconsistencies in his statements. It is surely true that one should think carefully and invest significant effort into the preparation of an important decision. Nonetheless, there is no point in "helping" people in their decision making through unnerving questions with the result that they actually lose all confidence in their own answers. Individuals will have much greater confidence in the validity of their statements if they are allowed to make statements about value intervals instead of exact point values. For clarification, compare the following two statements:

1. The combination of an annual salary of €100,000 and 25 days off is equal to an annual salary of €90,000 and 35 days off.
2. I prefer the combination of an annual salary of €100,000 and 25 days off to an annual salary of €90,000 and 30 days off, but do not prefer it to an annual salary of €90,000 and 40 days off.

The second statement can probably be made with lower mental costs and with more confidence in its validity than the first one.

There are two options for dealing with inconsistent information. First, you can generate unique information using mathematical techniques, e.g. by averaging the elicited values or applying mathematical/statistical methods of error minimization (Section 6.5.2). Second, as discussed in Section 5.3 with respect to incomplete value functions, you can try to manage the incomplete information by using intervals instead of exact values. To this end, we will discuss dominance tests (Section 6.5.3) and sensitivity analyses (Section 6.5.4).

6.5.2 Error minimization

The error minimizing approach interprets the indifference statements as random draws from a distribution around the true value. Let us assume there are n indifference statements that can each be denoted by an equation of the form $f_i(w_1, w_2, ..., w_m) = 0$. If the number of statements (n) is larger than the number of searched-for parameters (m = number of objectives), the corresponding system of equations will usually be over-determined. If we introduce an error term ε_i with $f_i(w_1, w_2, ..., w_m) = \varepsilon_i$ for every equation, the system is under-determined. It now seems natural to search for the objective weights w_r that – in absolute terms – lead to the lowest values of the error terms ε_i. This can be achieved, for instance, by means of a linear programming approach. We want to look at two possible procedures.

In the first option, you minimize the sum of the absolute values of the error terms. The objective function then is:

$$\text{Minimize} \sum |\varepsilon_i| \tag{6.7}$$

In a linear optimization model, this objective function can be expressed by the introduction of two positive variables ε_i^+ and ε_i^- with $\varepsilon_i = \varepsilon_i^+ - \varepsilon_i^-$. The objective function then becomes:

$$\text{Minimize} \sum (\varepsilon_i^+ + \varepsilon_i^-). \tag{6.8}$$

The second option minimizes the greatest absolute error. This is done by including an additional variable $\varepsilon_{\max}$ with $\varepsilon_i^+ \le \varepsilon_{\max}$ and $\varepsilon_i^- \le \varepsilon_{\max}$ in addition to the variables ε_i^+ and ε_i^-. The objective function then is:

$$\text{Minimize } \varepsilon_{\max} \tag{6.9}$$

Such an error-minimizing approach can always be applied when the preference information is given as a linear equation of the form

$$f_i(w_1, w_2, \ldots, w_m) = 0. \tag{6.10}$$

Let us return to our example using the trade-off method and let us assume that the decision maker would quote the following four trade-off statements instead of the necessary two.

$$\begin{aligned}
1&: (55{,}000, 40, *) \sim (45{,}000, 28, *)\\
2&: (55{,}000, 40, *) \sim (70{,}000, 52, *)\\
3&: (55{,}000, *, \text{good}) \sim (45{,}000, *, \text{excellent})\\
4&: (43{,}000, *, \text{good}) \sim (80{,}000, *, \text{bad}).
\end{aligned}$$

If we rearrange these four tradeoffs into appropriate equations, we obtain

$$\begin{aligned}
1&: w_1(v_1(55{,}000) - v_1(45{,}000)) + w_2(v_2(40) - v_2(28)) = 0\\
2&: w_1(v_1(55{,}000) - v_1(70{,}000)) + w_2(v_2(40) - v_2(52)) = 0\\
3&: w_1(v_1(55{,}000) - v_1(45{,}000)) + w_3(v_3(\text{good}) - v_3(\text{excellent})) = 0\\
4&: w_1(v_1(43{,}000) - v_1(80{,}000)) + w_3(v_3(\text{good}) - v_3(\text{bad})) = 0.
\end{aligned}$$

If we now compute the differences and insert the error terms, we obtain as constraints for the linear optimization approach:

$$\begin{aligned}
0.212w_1 - 0.3w_2 &= \varepsilon_1^+ - \varepsilon_1^-\\
-0.209w_1 + 0.3w_2 &= \varepsilon_2^+ - \varepsilon_2^-\\
0.212w_1 - 0.3w_3 &= \varepsilon_3^+ - \varepsilon_3^-\\
-0.563w_1 + 0.7w_3 &= \varepsilon_4^+ - \varepsilon_4^-.
\end{aligned}$$

Further constraints are:

$$w_l + w_2 + w_3 = 1$$

and

$$\begin{aligned}
\varepsilon_i^+, \varepsilon_i^- &\ge 0 \text{ for all } i = 1, 2, 3, 4\\
w_r &\ge 0 \text{ for all } r = 1, 2, 3.
\end{aligned}$$

If we now consider the objective function

$$\text{Minimize} \sum (\varepsilon_i^+ + \varepsilon_i^-),$$

we obtain the objective weights $w_1 = 0.40$, $w_2 = 0.28$, and $w_3 = 0.32$ as well as a value of the objective function of 0.0136. On the basis of these weights we obtain a value of 0.62 for the alternative consultancy, 0.70 for the university, and 0.28 for the sailing academy.

6.5.3 Dominance test

By the incomplete preference information, the set of possible attribute weights (or better: weight combinations) w_r is narrowed down. For every admissible weight combination and the resulting multiattribute value function we can calculate a preference ordering for the given alternatives. In the case in which we obtain the same preference ordering (for instance if we just compare two alternatives) for all admissible weight combinations, we can make unambiguous statements about the preference ordering of alternatives without definitely knowing the weights. If an alternative features the highest value of v over all admissible weights, it has to be optimal.

Let $V(I)$ be the set of value functions v that are compatible with the preference information I. We can then define a relation $\succeq_{V(I)}$ with

$$a \succeq_{V(I)} b \Leftrightarrow v(a) \geq v(b) \quad \text{for all } v \in V(I) \tag{6.11}$$
$$\text{and } v(a) > v(b) \text{ for at least one } v \in V(I),$$
$$a, b \in A \quad (\text{analogous for } \sim, \prec).$$

How can we tell if $a \succeq_{V(I)} b$ holds or not? Obviously, this relation does not hold if there is at least one value function v of $V(I)$ for which $v(a) < v(b)$ applies. Whether such a function exists can be determined quite easily via an optimization approach. It reads:

$$\text{Minimize } v(a) - v(b) \tag{6.12}$$

under the constraint $v \in V(I)$. If a negative minimum is produced by the objective function, a function v has to exist with $v(a) < v(b)$. It is hence possible that a is worse than b. If the optimization leads to a positive number, the statement $v(a) > v(b)$ holds for all functions v. From this fact, it follows that $a \succeq_{V(I)} b$, i.e. a is definitely better than b.

If the minimization of the objective function leads to a negative number, it can by no means be concluded that $b \succeq_{V(I)} a$ holds. In order to test this, we must use the objective function

$$\text{Maximize } v(a) - v(b). \tag{6.13}$$

If we obtain a negative maximum, $b \succeq_{V(I)} a$ holds. Again, a positive maximum leads only to the statement that $b \succeq_{V(I)} a$ does not hold.

If $\min[v(a) - v(b)] > 0$, a is dominating. If $\max[v(a) - v(b)] < 0$, b is dominating. If both the minimum and the maximum are zero, then a and b are of equal value for any admissible value function. Table 6-7 gives a summary of the potential cases.

Table 6-7: Results of the dominance test between two alternatives a and b

		$\max[v(a) - v(b)]$ < 0	$= 0$	> 0
$\min[v(a) - v(b)]$	< 0	b dominates a	b dominates a	no statement
	$= 0$	impossible	indifference	a dominates b
	> 0	impossile	impossible	a dominates b

This approach shows a way of narrowing down the number of relevant alternatives or of identifying efficient alternatives on the basis of a low level of information.

Under particularly favorable conditions, only one efficient alternative remains; in that case, this alternative is inevitably optimal. Figure 6-4 summarizes the approach.

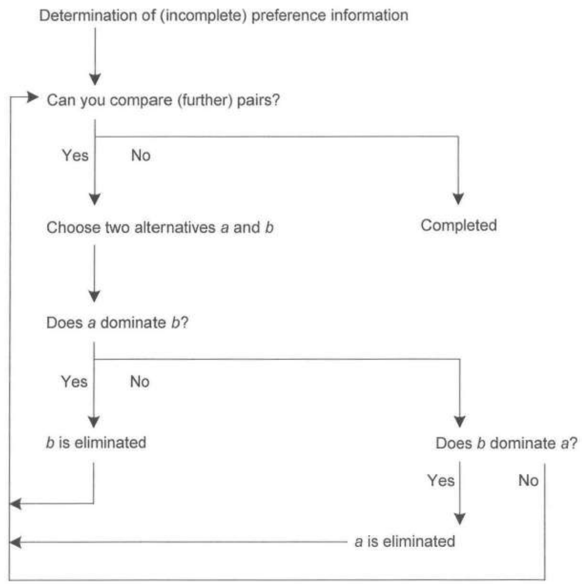


Figure 6-4: Procedure for the determination of dominance relationships under incomplete preference information

We will apply this approach to our example. In the application of the trade-off method, we had used in (6.4) the decision maker's indifference statement (€55,000, 60 hours, *) ~ (€30,000, 20 hours, *) to derive the weight relation $w_2/w_1 = 0.7$. Let us now assume that the decision maker is not willing to specify his preference more exactly than by the statement:

$$(€53{,}000, 60 \text{ hours}, *) \prec (€30{,}000, 20 \text{ hours}, *) \prec (€57{,}000, 60 \text{ hours}, *)$$

In this case, we cannot do any better than to define a range for the weight relation. Relying on the functional form of v_1 and proceeding in analogy to (6.4), we derive:

$$0.66 < w_2 / w_1 < 0.73 \tag{6.14}$$

Let us next assume that the decision maker is no longer willing to make the strong indifference statement

$$(€70{,}000, *, \text{bad}) \sim (€30{,}000, *, \text{excellent})$$

that we have seen in (6.5) and have used to derive the weight relation:

$$w_3 / w_1 = 0.91$$

Instead, he restricts the information to the weaker preference statement:

$$(€65{,}000, *, \text{bad}) \prec (€30{,}000, *, \text{excellent}) \prec (€75{,}000, *, \text{bad}).$$

We can use this information to derive a range for another weight relation:

$$0.85 < w_3 / w_1 < 0.96. \qquad (6.15)$$

Now we make use of the fact that weights in any case have only a relative meaning. For the resulting decisions, it does not matter whether we use the common normalization condition $w_1 + w_2 + w_3 = 1$ to fix the weights at some absolute level or set $w_1 = 1$, for instance, to achieve the same goal. This slightly different type of normalization will simplify the following calculations; if you feel more comfortable in the meantime with the usual type of normalization, we can always return to the $w_1 + w_2 + w_3 = 1$ condition at the end.

The interval statement then leads to the following conditions on the ranges of the weights:

$$w_1 = 1, \quad w_2 \in [0.66; 0.73] \quad w_3 \in [0.85; 0.96].$$

We now do not need to care about any weight sum when choosing the other weights.

We want to check whether we can make a decision despite the incomplete information, i.e. whether there are dominance relations. Table 6-8 summarizes the attribute evaluations of the given alternatives once again.

Table 6-8: Attribute evaluations of the alternatives

Alternative	$v_1(x_1)$	$v_2(x_2)$	$v_3(x_3)$
a	1	0	0.7
b	0.6	0.5	1
c	0	1	0

We will start by screening alternatives a and b for a dominance relation.

a* vs. *b

To save time and effort, it is always recommendable to begin a dominance test (at least if it is done manually) by determining the preferred alternative for one of the admissible weight combinations. You can thereby immediately eliminate one of the two possible directions of dominance. In our example, this is particularly simple and obvious. From Section 6.4.2, we already know that the indifference statements that were elicited via the trade-off method (and that are placed within the broader ranges that we consider now) led to a preference for alternative b over alternative a. The only possible dominance direction is thus a dominance of b over a.

To analyze this possibility, we must solve the problem

$$\begin{gathered}\text{Maximize}\\ v(a) - v(b)\\ = w_1(v(a_1) - v(b_1)) + w_2(v(a_2) - v(b_2)) + w_3(v(a_3) - v(b_3))\\ = w_1(1 - 0{,}6) + w_2(0 - 0{,}5) + w_3(0{,}7 - 1)\\ = 0{,}4w_1 - 0{,}5w_2 - 0{,}3w_3\end{gathered}$$

to see whether a positive value difference $v(a) - v(b)$ can occur for the given weight ranges. To maximize this expression, all weights with a positive coefficient have to be chosen at the upper limit of the range while all weights with a negative coefficient have to be chosen at the lower range limit. Remember that the weight w_1 was set to 1 in advance to fix the absolute level of the weights. We thus do not have to consider any further normalization constraints when choosing w_2 and w_3. In our specific case, we choose the smallest admissible figures for the weights w_2 and w_3 (0.66 and 0.85) to obtain the maximum. We get:

$$0.4w_1 - 0.5w_2 - 0.3w_3 = 0.4 \cdot 1 - 0.5 \cdot 0.66 - 0.3 \cdot 0.85 = -0.185.$$

Hence, even if we make optimal use of the flexibility in choosing the weights w_2 and w_3 in favor of alternative a, a negative value difference (i.e. a lower evaluation of a compared to b) will occur. Alternative a is thus dominated by alternative b and can be eliminated. In case you still have doubts about this result because we have not used the standard type of normalization, simply check what would change if the weights $w_1 = 1$, $w_2 = 0.66$, and $w_3 = 0.85$ that we have chosen for maximization were divided by their sum, namely 2.51. You will realize that the sum of the thus adjusted weights is 1, the restrictions concerning the weight relation w_2/w_1 and w_3/w_1 are still satisfied and positioned at the lower limit of the admissible range, and that the sign of the maximal value difference is still negative (as a matter of fact, it is simply scaled down by the factor 2.51). This thought experiment illustrates once again that attribute weights have a meaning only in a relative sense; a scaling of the complete set of weights does not change the resulting preferences.

b* vs. *c

It remains to check whether we have a situation of dominance between b and c. We start again by exploiting our knowledge from Section 6.4.2 to conclude that a dominance of alternative c over alternative b can be excluded immediately. Let us thus check whether there is a dominance of b over c:

$$\begin{gathered}\text{Minimize}\\ v(b) - v(c)\\ = w_1(v(b_1) - v(c_1)) + w_2(v(b_2) - v(c_2)) + w_3(v(b_3) - v(c_3))\end{gathered}$$

$$= w_1(0{,}6-0)+w_2(0{,}5-1)+w_3(1-0)$$
$$=0{,}6w_1-0{,}5w_2+1w_3.$$

To minimize the expression, we have to look at the coefficients again. The weight w_2 has a negative coefficient and is thus chosen as large as possible ($w_2 = 0.73$). The weight w_3 has a positive coefficient and is thus chosen to be as small as possible ($w_3 = 0.85$); the weight w_1 was normalized to the level 1 anyway. We find that even if we choose all the weights in favor of alternative c, a positive value difference of 1.085 remains. Hence, alternative b also dominates alternative c.

We have seen that we were able to clearly identify an optimal alternative despite the incomplete information about the weights. Of course, this is not always the case. The narrower the intervals are, the better the chances of identifying an optimal alternative become. If the intervals are very large, it might not even be possible to identify and eliminate any dominated alternative at all. In this case, the decision maker has to increase his cognitive effort in order to produce smaller intervals.

In our opinion, the approach "search for a solution on the basis of incomplete information first and only increase the effort if necessary" is extremely important for practical applications. It would be uneconomical to determine precise attribute weights with a lot of effort if a decision was as well reachable with much lower precision. Moreover, the willingness of the decision maker to subject himself to the strenuous weight determination process will probably decline with increasing effort. The decision maker will not trust the very precise weight determination because he realizes that the accuracy is just imagined. In contrast, if he is allowed to specify intervals that he feels comfortable with, he will rightly trust the optimal solution (if one can be determined with the incomplete information). For additional methods for incomplete weight determination we refer to Weber (1983), Kirkwood and Sarin (1985), and Weber (1985).

6.5.4 Sensitivity analyses of the weights

In all cases in which specific results (here: values of the alternatives) are derived on the basis of exact parameters (here: weights), you should think about whether the result might change if you altered one or more of the parameters slightly. For example, the aforementioned graduate student might have identified certain attribute weights and thereby arrived at a decision in favor of the position at the university. However, he is well aware of the fact that these weights are not completely accurate and might vary depending on daytime, mental state, or other parameters. It is therefore important to know if his decision in favor of the university is quite stable or may be quickly reversed if the weights are slightly changed, for instance in favor of the consultancy. The consultancy has an advantage over the university in terms of the attribute initial salary. A higher weight attached to the initial salary could thus result in the consultancy being valued higher than the university.

Sensitivity analyses of weights are not completely trivial. You cannot simply determine the impact of an isolated increase of the salary weight w_1 as long as the normalization condition $\Sigma w_r = 1$ still has to hold; you thus need assumptions on how the other weights change if the weight of one attribute is altered.

One possible assumption is that if you increase the weight of the initial salary, the weights of all other attributes are reduced in such a way that their respective relations remain unchanged. Suppose that the decision maker in our example originally assumed weights of $w_1 = 0.40$, $w_2 = 0.28$, and $w_3 = 0.32$. He wants to vary w_1, while leaving w_2/w_3 at the original ratio of 0.28/0.32. Formally, the following conditions have to be satisfied:

$$w_1 \in [0;1]$$
$$\frac{w_2}{w_3} = \frac{0.28}{0.32}$$
$$w_1 + w_2 + w_3 = 1$$

hence

$$w_2 = (1-w_1)\cdot\frac{0.28}{(0.28+0.32)}$$
$$w_3 = (1-w_1)\cdot\frac{0.32}{(0.28+0.32)}.$$

In the upper part of Figure 6-5 the weighting functions w_2 and w_3 are displayed as a function of w_1. Consequently, the values of the three alternatives

$$v(a) = 1\,w_1 + 0\,w_2 + 0.7\,w_3$$
$$v(b) = 0.6\,w_1 + 0.5\,w_2 + 1\,w_3$$
$$v(c) = 0\,w_1 + 1\,w_2 + 0\,w_3$$

follow as seen in Figure 6-5.

This sensitivity analysis shows that w_1 needs to rise to a level of 0.50 to change the decision outcome in favor of the consultancy. The decision maker should contemplate if such a weight shift might be realistic and possible. To do so, he could create two hypothetical alternatives that have the same evaluation according to the new weights $w_1 = 0.50$, $w_2 = 0.23$, and $w_3 = 0.27$, and ask himself if there is any way that he would consider these alternatives equally attractive.

Sensitivity analyses and dominance tests in situations of incomplete information (as considered in Section 6.5.3) are strongly related from a conceptual point of view. In the former case, we start at a given weight combination and ask ourselves how much imprecision of the weights would be acceptable without changing the decision outcome. In the latter case, we start with imprecise weights and ask ourselves whether this imprecision is sufficiently marginal to arrive at the same decision outcome for all admissible weight combinations.

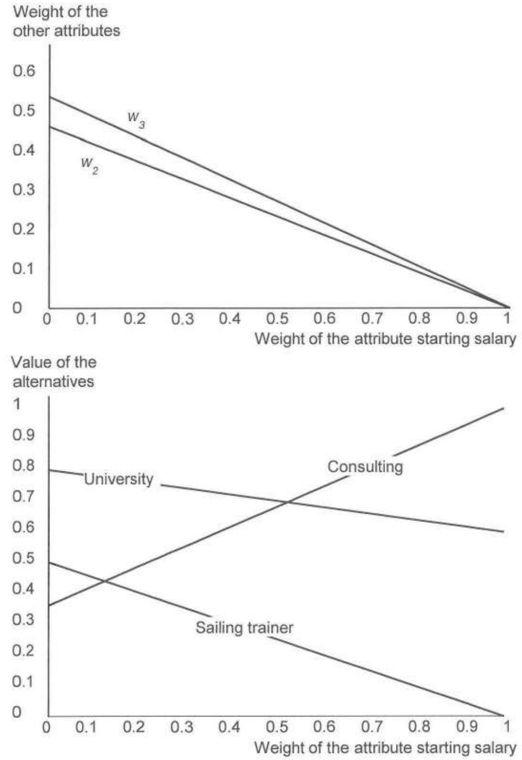


Figure 6-5: Sensitivity analysis for the weight of the attribute initial salary, w_1

6.6 The dependence of the weights on the attribute range

The weights w_r in an additive model depend on the intervals $[x_r^-, x_r^+]$ over which the attribute value functions v_r are defined. If the interval is narrow, the weights that are attached to it have to be smaller compared to a broader interval; this is due to the fact that for the small interval, the value difference between the worst and best attribute levels is larger than the value difference that we obtain for the same attribute levels in the extended interval (with a renormalized value function). You have to understand that attributes per se do not have any weights attached to them. "Health is more important to me than my salary" is not a statement that is of any

relevance for multiattribute evaluations. This insight is frequently disregarded, for instance when corporate objectives have to be ranked according to their importance in surveys, etc. (see e.g. R. L. Keeney 2002). "The satisfaction of our employees is equally important to us as is the realization of profits" does not convey any information. Another example: in a test of vacuum cleaners, the attribute "safety" receives the weight factor 0.15, the attribute "handling" the weight factor 0.25, and so on. Does this really mean that handling is more important than safety? Without knowing the intervals of the attributes, the weights are meaningless. If the worst possible level of the attribute safety is attached to a device that delivers a deadly power surge every time it is turned on, any sensible consumer would obviously attach a weight of 1 to the attribute safety and 0 to any other attribute. However, if the attribute level interval ranges from "low risk of injury if opened improperly" to "no danger whatsoever", the attribute safety does not need to be weighted so heavily.

Let us assume we had an additive model with the weights w_r and the value functions v_r that were determined over the intervals

$$B_r = [x_r^-, x_r^+].$$

The value difference between the best and the worst level of the r-th attribute

$$\Delta v_r(B_r) = v_r(x_r^+) - v_r(x_r^-) \tag{6.16}$$

equals 1 as agreed upon before. If you consider a broader interval $B'_r = [x_r^{-\prime}, x_r^{+\prime}]$ and thus "extend" the value function accordingly, the value differences between the alternatives in attribute X_r shrink after normalization by a ratio of $\Delta v_r(B_r) / \Delta v_r(B'_r) < 1$. Accordingly, the normalized value differences grow if you proceed to a value function with a narrower interval. It is important to recognize that the change of the value difference is not driven by altered preferences of the decision maker with respect to these attribute levels; instead, it is simply a consequence of a different normalization. In order to compensate for this change in value difference, you would have to multiply the attribute weight with $M = \Delta v_r(B'_r) / \Delta v_r(B_r) = \Delta v_r(B'_r)$. The resulting weight combination would again correctly represent the decision maker's preference. The minor problem remains that the sum of the adjusted weights is no longer equal to 1. However, this problem can be easily solved in the standardized way: all weights are divided by the sum of the weights and we obtain:

$$w'_r = \frac{Mw_r}{\sum_{i \neq r} w_i + Mw_r}. \tag{6.17}$$

and for the other attributes X_j:

$$w'_j = \frac{w_j}{\sum_{i \neq r} w_i + M w_r}. \tag{6.18}$$

We want to explain this important connection by means of a simple example: Two used cars are assessed and the evaluation is solely based on the attributes sales price (X_1) and mileage (X_2);both value functions are assumed to be linear. For the price, the value function is determined over the interval [€10,000, €20,000] and for the mileage, it is determined over the interval [20,000km, 80,000km]. The smaller level is obviously the better one in both cases and hence receives the value 1. The larger number is worse and is assigned the value 0. The individual value functions are thus

$$v_1(x_1) = 2 - \frac{1}{10{,}000} x_1$$
$$v_2(x_2) = \frac{4}{3} - \frac{1}{60{,}000} x_2.$$

Weights of $w_1 = 0.6$ and $w_2 = 0.4$ may be assumed to apply. As can be seen in Table 6-9, the used car a receives a higher total value under these assumptions and is hence chosen over b.

Table 6-9: Evaluation of two used cars

Used car	Price $w_1=0.6$	v_1	Mileage $w_2=0.4$	v_2	Total value
a	€10,000	1	50,000km	0.5	0.8
b	€15,000	0.5	20,000km	1	0.7

If you choose a different interval for the determination of the value function over the attribute sales price, e.g. [€5,000, €25,000], you obtain (under the assumption that the value function stays linear over a broader interval)

$$v'_1(x_1) = \frac{5}{4} - \frac{1}{20{,}000} x_1.$$

If we use this function and maintain the weights 0.6 and 0.4, we come to an evaluation as presented in Table 6-10 and realize that car b is now preferred. Although the difference between the attribute levels is still the same, the influence of the sales price was reduced due to the adjusted (lowered) value difference. For that reason, the ordering of the alternatives is reversed.

To obtain a consistent evaluation, we need to adjust the attribute weights as well. If we insert the new lower or upper limit in the old value function, we obtain $v_1(5{,}000) = 1.5$ or $v_1(25{,}000) = -0.5$, respectively, resulting in a value difference of $\Delta v_r(B'_r)$ of 2 and thus also an adjustment factor of $M = 2$.

Table 6-10: Evaluation of two used cars under an increased range of the value function for the sales price

Used car	Price $w_1=0.6$	v_1	Mileage $w_2=0.4$	v_2	Total value
a	€10,000	0.75	50,000km	0.5	0.65
b	€15,000	0.5	20,000km	1	0.7

The correct new weights are then calculated as

$$w'_1 = \frac{0.6 \cdot 2}{(0.6 \cdot 2 + 0.4)} = 0.75$$
$$w'_2 = \frac{0.4}{(0.4 + 0.6 \cdot 2)} = 0.25,$$

and their application leads to $v(a) = 0.6875$ and $v(b) = 0.625$, hence yielding a preference for car a.

Any method for the determination of weight factors that does not refer to particular attribute intervals is doomed to fail. As indicated before, this includes the direct-ratio method – unless the decision maker knows the attribute intervals and is well aware of them when making statements concerning the "importance" of the attributes.

6.7 Cognitive biases in the determination of the weights

6.7.1 The range effect

As we have explained in the previous section, weights can only be interpreted in a meaningful way with respect to predefined intervals of the objective variable. An approach that tries to derive weights from general statements about the importance of the attributes without considering any intervals cannot yield credible results. A method like the direct-ratio could be "salvaged" if the assumed intervals were made transparent before the decision maker is asked about the "importance" of the attributes. In such a case, the decision maker could ensure that his statements about the importance vary accordingly with the intervals in the theoretically correct proportion. For a larger interval, the decision maker would hence increase the importance and decrease it for a smaller interval. However, in the experimental studies of von Nitzsch and Weber (1991), it became apparent that decision makers often did not adjust their importance statements at all when faced with different intervals and on average did so only insufficiently. Against the background of these insights, the recommendation to avoid the direct-ratio method has to be further emphasized.

6.7.2 The splitting effect

During the process of generating a system of objectives, it can prove useful to break down an objective into subordinate objectives if doing so helps to clarify the importance of the objective and improves its measurability. Obviously, however, the splitting should not lead to different weights. If the objective "career perspectives" in our job example has a weight of 0.32, then the sum of the weights of two subordinate objectives, say "chances for a responsible position" and "salary increase", should be the exact same (0.32) after the splitting.

Experimental studies show that, in contrast to this claim, people tend to overweight an objective that is broken down into subordinate objectives when compared to the same objective that is not split up (Eisenführ and Weber 1986; Weber et al. 1987). This means that if the second of two objectives X and Y is split up into two subordinate objectives Y_1 and Y_2, then the weight of X decreases, whereas Y_1 and Y_2 jointly obtain a higher weight than Y before.

This phenomenon leads to the problem that disaggregated objectives are overweighted and aggregated objectives underweighted. How is it possible to correct for such a bias? For one thing, it is certainly helpful to be aware of the bias. Additionally, weights for different aggregations can be determined and then checked for consistency. If it turns out that a disaggregated objective has more weight attached to it than its aggregated counterpart, you could confront the decision maker with this fact whereupon he might alter his assessment towards more consistency. If he is not able or willing to do so, then we have no means with which to determine the "correct" weights. However, we could try to go ahead with the incomplete information as described in Section 6.5. Therefore, let w_r* be the measured weight of the aggregate objective and w_{ri}* be the measured weights of the subordinate objectives satisfying $\Sigma w_{ri}* > w_r*$. We could then check if we can identify individual dominance relationships in pair wise comparisons between alternatives for weights w_{ri} subject to the constraints

$$\begin{aligned} w_{ri} &\le w_{ri}^* \quad \text{for all } i \\ \sum_i w_{ri} &\ge w_r^*. \end{aligned} \tag{6.19}$$

In a particularly favorable case, we might even be able to end up with a single optimal alternative.

For a discussion of cognitive biases in the process of determining objective weights, we refer to Borcherding et al. (1991) as well as Weber and Borcherding (1993).

Questions and exercises

6.1

Harry is looking for people in his circle of friends to accompany him on a sailing trip for a few weeks. It is an important problem as he knows how annoying the

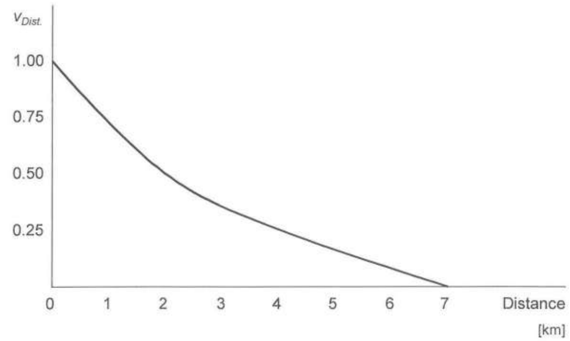
wrong people can be on such a trip. He hence makes the following list of important attributes which he would prefer in the candidates:

- reliable
- healthy
- good humor
- sporty
- experienced with sailing
- sociable
- technical understanding
- not egoistic

Test for mutual preferential independence between the attributes and if it is given also for difference independence.

6.2

Ernie Eternity is looking for a new flat at his new place of study. He has the two objectives "rent as low as possible" and "as close to the university as possible". The offers vary between €200 and €600 for rental fees and between 0km and 7km in terms of distances. His value function for the rent is linear; the other value function is given in the following diagram.



(a) What could a possible explanation for the specific shape of the value function for distance be?
(b) Ernie thinks that a flat 4km away from the university costing €300 is as attractive as a flat 2km away from the university costing €500. Use this statement to calculate the two attribute weights w_1 and w_2 in an additive value function.
(c) What price would be acceptable to Ernie for a flat that is 0km away from the university to remain preferable to a flat 2km from the university at a cost of €200?

6.3
Manager Freddy Front wants to buy a notebook for his frequent travels. He considers five alternatives which are all on the same technical level. He wants his choice to depend on the two attributes "battery life" and "weight" only.

(a) The manager goes to a specialist store to collect the relevant information. He tells the shop assistant: "Low weight of the notebook is more important than the battery life". Please comment on this statement.
(b) He intends to solve the problem by using an additive value function. For this purpose, Freddy generates attribute value functions
  - for the battery life between 0 and 8 hours,
  - for the weight between 0 and 6kg.
(c) Draw plausible graphs for the two value functions and explain them.
(d) Assume the attribute value functions are linear. The worst level of the attributes is assigned an attribute value of zero and the best level is assigned a value of 1 for both attributes. Then, the weights for the two attributes have to be defined. Freddy thinks that a notebook of 2kg with a battery lasting 4 hours is as good as a laptop of 3kg with a battery lasting 6 hours. What are the weights that can be derived from this statement?

6.4
Having successfully graduated from her business administration program, Steffi Starter cares about three attributes when selecting between different job offers: "starting salary", "distance from home", and "expected weekly workload". She is sure that those three characteristics are difference independent. The table presents the offers that she has selected during her first interview trip.

Job offer	Starting salary in €1,000 p.a.	Distance from home in km	Work load in hours/week
a	80	40	60
b	40	20	30
c	50	0	50
d	70	50	55

The value functions are linear within the intervals [€40,000/year, €80,000/year], [0km, 50km], and [30h/week, 60h/week]. Steffi generates the following point evaluations when determining her value function:

(€40,000/year, 50km, 60 hours/week) = 0 points
(€40,000/ year, 0km, 60 hours/week) = 30 points
(€80,000/ year, 50km, 60 hours/week) = 70 points
(€40,000/ year, 50km, 30 hours/week) = 100 points.

(a) Which method did Steffi use?
(b) Which weights w_1, w_2, and w_3 can be derived and which offer is the best?

(c) Steffi is unsettled by the result and wants to attempt less exact evaluations. She is sure about the best and worst alternatives, but less so concerning the other two. Therefore, she states the following point intervals for the two hypothetical intermediate job offers:

(€40,000/year, 0km, 60 hours/week) = 20-30 points
(€80,000/year, 50km, 60 hours/week) = 60-80 points.

Is it still possible for Steffi Starter to conclude which is the best offer?

6.5
Engineering student Peter Singel is only interested in the criteria "revolutions per minute" and "load capacity in kg" when it comes to selecting a spin-dryer. He has linear increasing single value functions for both attributes. He defines his weights to be 0.6 for revolutions and 0.4 for load capacity in the ranges from 600 to 1,400 revolutions and from 3 to 5.5kg load capacity. When checking the current offers, he realizes that there are no spin-dryers with more than 1,200 revolutions per minute. Does this have an effect on his aggregate value function? If so, which?

6.6
Recreational diver Leo Prene wants a new watch for his birthday. He explains to his girlfriend what is important to him: (1) preferably low price, (2) preferably long guarantee period, and (3) preferably water resistant in deep water. He considers these attributes to be difference independent. After collecting information, he assumes the following intervals to be relevant: prices from €50 to €200, guarantee for 4 to 16 years, and water resistance at depths from 10 to 50 meters. He tells his girlfriend that his attribute value functions are:

$$v_1(x_1) = a + b / x_1$$
$$v_2(x_2) = c + d\sqrt{x_2}$$
$$v_3(x_3) = e + fx_3$$

(a) Determine the normalized attribute value functions in the interval [0; 1], i.e. determine the parameters *a* to *f*.
(b) What do Leo's answers to the following questions have to look like to be consistent?
  - To which amount does the price starting from €50 have to increase to generate half the value?
  - Which transition is more valuable – the one from 4 to 9 years or the one from 9 to 16 years?
(c) Leo informs his girlfriend about the following indifferences:

$$(€200, 16\text{ years}, 12\text{ meters}) \sim (€80, 4\text{ years}, 12\text{ meters})$$
$$(€70, 16\text{ years}, 10\text{ meters}) \sim (€70, 4\text{ years}, 35\text{ meters}).$$

What is Leo's multiattribute value function?

6.7

Mr. Miller wants to buy a new coffee maker. He only cares about the three attributes: price (how expensive), capacity (how many cups) and guarantee (how many months). At the Coffee-Maker-Palace, he finds a huge variety of coffee makers ranging in price from \$29 to \$529, in capacity from 2 to 12 cups, and in guarantee time from 0 to 24 months.

Mr. Miller made sure that his preferences with respect to these objectives fulfill the necessary independence conditions to justify the use of an additive model. He further determined that his attribute value functions are linear (decreasing in price and increasing in the other attributes).

(a) To determine his attribute weights, he uses the Swing method and comes up with the weights $w_1 = 0.4$ (for price), $w_2 = 0.5$ (for capacity) and $w_3 = 0.1$ (for guarantee time). Please describe in detail how Mr. Miller must have derived these weights (i.e. describe the application of the Swing method for the explicit scenario given above and make assumptions about Mr. Miller's preferences that are in accordance with the resulting weights).

(b) What would have changed in the process if Mr. Miller's attribute value functions were reflecting diminishing value sensitivity (i.e. were not linear)?

(c) For Mr. Miller's final choice, there remain two coffee makers: the Coffee-Max3000, featuring the attribute bundle (\$229, 8 cups, 12 months) and the CoMa-Galactica, featuring the attribute bundle (\$529, 12 cups, 24 months). Which coffee maker should Mr. Miller buy?

(d) The direct ratio method was criticized in this chapter to be a "flawed" procedure for determining the attribute weights. Shortly explain the procedure and the problem that results from the application of this method.

6.8

Larah recently finished high school and is making plans for her college education now. She is sure that business administration is the perfect major for her interests but is unsure about the optimal college town; she hence performs some systematic analysis.

There are three options she considers: the University of Kiel, the University of Muenster, and the University of Paderborn. She has identified three objectives that are important for her and made sure that these objectives satisfy all preference independence claims that allow her to use an additive model; these objectives are: distance from home (less is better), reputation of the school (more points in the latest ranking are better), and tuition and fees (less is better). She has collected all relevant information in the following table:

	Alternative *a*: Kiel	Alternative *b*: Münster	Alternative *c*: Paderborn
Attribute 1: distance from home	250km	150km	50km
Attribute 2: reputation of the school	60 points	80 points	30 points
Attribute 3: tuition and fees	€0	€275	€500

Larah uses the bisection method to determine her attribute value functions and determines that they are linear for all attributes.

(a) What was the first indifference statement that Larah has constructed in this process (for the attribute "distance from home")? What point of the value function did she determine by this step?

(b) Determine the valuations of all outcomes in the table.

(c) Larah determined the attribute weights to be $w_1 = 0.5$, $w_2 = 0.1$ and $w_3 = 0.4$. What alternative turns out to be the best for Larah given these weights?

(d) Larah has used the trade-off method to determine the attribute weights. On the slip of paper that she used for the procedure one finds a line that starts with:

$$(250\text{km}, 80\text{ points}, \bullet) \sim (\ldots\ldots\ldots$$

Complete the missing part of this line and explain why this is the correct completion.

(e) When measuring the distance from home again, Larah realizes that Kiel is farther away from home than she thought so far. She hence repeats the complete analysis (leaving everything else the same). Which of the following four weight-combinations (A-D) could be the result of the updated procedure? Explain shortly why this is the only possible combination of the four.

A: $(w_1 = 0.50, w_2 = 0.10, w_3 = 0.40)$
B: $(w_1 = 0.60, w_2 = 0.08, w_3 = 0.32)$
C: $(w_1 = 0.40, w_2 = 0.12, w_3 = 0.48)$
D: $(w_1 = 0.70, w_2 = 0.05, w_3 = 0.25)$

6.9

The decision concerning a camera purchase is intended to be made according to the four attributes price, image quality, equipment, and handling. The shortlist consists of three cameras. They meet the requirements as stated in the table:

	Price	Image quality	Equipment	Handling
Modell *A*	0	0.8	1	0.9
Modell *B*	0.2	1	0	1
Modell *C*	1	0	0.5	0

The buyer believes that the conditions of an additive model are fulfilled. For the given intervals, the attributes "image quality", "equipment", and "handling" are equally important to him; he however does not know how to determine the weight for price relative to the other weights.

Conduct a sensitivity analysis that shows which camera is optimal in which weight ranges for price.

6.10

Four attributes are relevant for the assessment of a certain device: function (X_1), safety (X_2), sound (X_3), and price (X_4). The shortlist consists of the three models *a*, *b*, and *c*. The team of experts that is dealing with the buying decision defines the weights as $w_1 = 0.4$, $w_2 = 0.25$, $w_3 = 0.15$, and $w_4 = 0.2$ for the underlying attribute domains.

The experts conclude after some discussion that the attribute "function" is too global to be clearly evaluated with respect to the devices; they thus split the attribute into three lower-level attributes X_{11}, X_{12}, and X_{13}. After a new weighting procedure, the results are:

$w_{11} = 0.25$ $w_2 = 0.20$
$w_{12} = 0.12$ $w_3 = 0.11$
$w_{13} = 0.16$ $w_4 = 0.16$.

The experts are not able to eliminate the inconsistency between the two weightings and hence decide to define as valid all weights between the two measurements in order to examine whether one or even two of the three alternatives are dominant.

The table depicts the evaluations of the alternatives for the six attributes:

	X_{11}	X_{12}	X_{13}	X_2	X_3	X_4
a	1	0.7	0	1	0	0.3
b	0	0	0.8	0.6	1	1
c	0.6	1	1	0	0.8	0

Sketch the optimization approach that tests whether alternative *a* or *b* is dominant, given this data. Try to find out (by trial and error) whether you can detect if one alternative is dominated.

Casy Study 1: Safety standards for oil tankers

Source: Ulvila und Snider (1980), p. 81-96.

Problem

After a series of tanker accidents at the end of 1976 and the beginning of 1977, U.S. President Carter announced in a message to Congress in March 1977 measures against the problem of marine pollution by oil, including an international conference to discuss an amendment to the existing agreements. Within a month, the IMCO (Intergovernmental Maritime Consultative Organization), an agency of the United Nations, consented to convene an international conference on tanker safety, which was to be held in February 1978. Since the usual preparation time of such conferences amounts to 4-5 years, accelerated procedures had to be used to assess the proposals of the U.S. and other countries.

The U.S. Coast Guard, responsible for representing the U.S. in conferences of this kind, began immediately with the preparations. Due to the short time available, it was decided that a quantitative model should be developed with which the proposals could be assessed. Therefore, experts of a consultancy intended to contribute the theoretical decision analysis know-how were included in the project group.

Procedure

The project group followed six steps:

1. Identification of the objectives relevant for the assessment made by different countries (not only the U.S.) about standards regarding tanker safety in terms of construction and equipment.
2. Definition of alternative standards and assessment of their attribute levels on the different objectives.
3. Check for a possible dominant alternative.
4. As there was no dominant alternative, weights for the different objectives were determined. Then, the attractiveness of various standards was assessed from the view of the U.S. and other countries.
5. Sensitivity analyses to check the results' stability against the assumptions.
6. Utilization of the model to search for new, enhanced alternatives.

The objectives

The team identified 11 objectives. The objectives as well as the attributes and the intervals for the attributes' levels are summarized in the following table. Attributes' levels are not measured in absolute terms but rather in differences from the "status quo", which would have presumably happened if the conference were to end without a result.

No.	Objective	Attribute	Level interval
1	lowest possible oil outflow from tankers worldwide	oil discharged into the oceans in tons per year	reduction of 0-28%
2	lowest possible oil outflow in U.S. territory	oil discharged within 50 miles of U.S. coastline in tons per year	reduction of 0-50%
3	maximal safety	reduction in injuries/fatalities and property damage	reduction of 0-3 fires/explosions and 20 fatal casualties per year
4	minimal costs	costs of implementation in USD	0-10bn USD
5	minimal costs per ton of circumvented oil outflow	costs (attribute 4) divided by outflow (attribute 1)	105-2,560 USD per ton
6	ease of passing cost to costumers	cost increase for oil	0-2%
7	minimal effects on existing charter agreements		*
8	maximal tanker capacity	point in time where supply and demand are balanced	1980-1982
9	maximal effects on utilized capacity of shipyards	orders for new ships as well as conversions	*
10	maximal competitive advantage due to modernized tankers	decommissioning of old tankers	0-60%
11	maximal enforceability	extent to which enforcement of the standard relies on equipment rather than human performance	*

* no quantitative indication

Since an additive model was to be used, those attributes exhibiting preferential dependencies were redefined until an (approximately) additive model was established. For instance, there was a dependency among the costs and the reduction of outflow. Some experts stated that costs do not matter much for them as long as the outflow would be significantly reduced; they would not, however, spend money for nothing. Therefore, the fifth criterion (costs per ton) was implemented.

The alternatives and their valuation

In a couple of meetings of the IMCO in the year 1977, three proposals were made, called "U.S. proposal", "package 1", and "package 2". A fourth alternative would be to stay with the status quo. For each alternative and objective, a point score was determined. The worst level got assigned 0 points while the best level got assigned 100 points. U.S. administration officials agreed with the project team on the assessment of the different proposals on the attribute level. It was assumed that the affected countries and other parties would differ only in their objective weights. Since there was no dominant alternative, the objective weights had to be derived. The swing-method was used for that. For instance, the following question had to be answered: What is more important - a reduction in costs of 10bn USD or the circumvention of three tanker explosions with 20 fatal casualties? The relative weightings regarding changes in the attribute levels were determined and normalized, yielding a sum of one. Several consistency checks were conducted to obtain unambiguous weights.

Experts of the U.S. coast guard calculated a weighting scheme not only for the U.S. but also for 21 of the 60 participating countries. By multiplying the attribute scores with the individual weights and subsequently summing up, a total score per proposal and per country was derived, allowing the ordering of the proposals as well as the differences in terms of value to be estimated. By doing so, possible courses of negotiations were anticipated. (In previous conferences, the U.S. delegation was sometimes surprised by other delegations' viewpoints.)

Results

The model supplied possible differences in viewpoints as well as their origins which had to be expected for the conference. Furthermore, it also demonstrated a possibility of generating new proposals which might more probably be the basis for a consensus compared to the original packages. Indeed, a new package was defined for which a higher degree of consensus based on the individual countries' value functions was expected. Interestingly, this new package came close to the final negotiation result. The conference ended with the successful passing of significant measures to improve tanker safety and protection against oil catastrophes.

Case Study 2: Choosing and office location through even swaps

Source: Hammond, Keeney und Raiffa (1998), p. 137-149.

The technical consultant Alan Miller was required to make a decision concerning his workplace since the lease of his current office was about to expire. Therefore, he defined five objectives that he needed the new location to fulfill: a short commute from home, good access to his clients (measured by the percentage of customers whose office is within an hour's lunch-time drive), good office services (such as clerical assistance, copiers and fax machines, and mail service), sufficient space, and low costs. Additionally, he clarified for himself that he assumes mutual preference independence for the attributes. After scanning the papers, Alan surveyed numerous places and, eliminating those which clearly did not meet his standards, shortlisted four promising alternatives: Parkway, Lombard, Baranov, and Montana. In order to structure the decision, he developed a consequences table which contrasted the alternatives in terms of their consequences for each of the objectives. However, it was impossible to conduct an overall appraisal per alternative as the variety of objectives required different measurement systems, constraining direct comparability. The even-swaps-method should thus provide a way to make a trade-off among the alternatives. Alan's consequences table should serve as a basis for this purpose:

	Parkway	Lombard	Baranov	Montana
commute in minutes	45	25	20	25
customer access within one hour (%)	50	80	70	85
office services (qualitatively assessed)	A	B	C	A
office size (square feet)	800	700	500	950
monthly cost ($)	1,850	1,700	1,500	1,900

Following the method of even swaps, in a next step, Alan sought to eliminate alternatives by the criterion of dominance. Despite no dominance among the alternatives, he found that Parkway was at least almost dominated by Montana which was disadvantageous only in terms of monthly cost. Hence, a first even swap was conducted to increase comparability between those two alternatives and to maybe create a dominance situation. Increasing the monthly cost of Parkway by $50 onto the level of Montana could, according to Alan's subjective utility appraisal, be compensated by reducing commute time by ten minutes. The swap revealed that Parkway was completely dominated by Montana now und could thus be eliminated:

	~~Parkway~~	Lombard	Baranov	Montana
commute in minutes	~~45~~ 35	25	20	25
customer access within one hour (%)	50	80	70	85
office services (qualitatively assessed)	A	B	C	A
office size (square feet)	800	700	500	950
monthly cost ($)	~~1,850~~ 1,900	1,700	1,500	1,900

Having reduced the number of alternatives to three, none of them dominating any other, a further series of even swaps had to be subsequently conducted to constrain the number of objectives and to initiate further eliminations. Scanning the table, Alan noticed similar values in commute time for all three alternatives. If Baranov's 20-minute commute was increased to 25 minutes by an even swap, an equal value of 25 over all alternatives would result, and therefore the objective could be neglected in further considerations. Alan assessed an 8 percentage point increase in customer access as a suitable compensation for the change in commute and undertook the swap, crossing out the objective of commuting time:

	Lombard	Baranov	Montana
~~commute in minutes~~	~~25~~	~~20 25~~	~~25~~
customer access within one hour (%)	80	~~70~~ 78	85
office services (qualitatively assessed)	B	C	A
office size (square feet)	700	500	950
monthly cost ($)	1,700	1,500	1,900

The next objective to be eliminated was office services. Setting level B as a standard, two even swaps were made with the objective of monthly cost. By increasing monthly costs by $200 for Baranov and decreasing them by $100 for Montana, necessary shifts in office services were compensated. With that objective being no longer crucial for the decision, Alan identified a dominance of Lombard over Baranov and, hence, crossed out Baranov:

	Lombard	~~Baranov~~	~~Montana~~
customer access within one hour (%)	80	78	85
~~office services (qualitatively assessed)~~	B	~~C~~ B	~~A~~ B
office size (square feet)	700	500	950
monthly cost ($)	1,700	~~1,500~~ 1,700	~~1,900~~ 1,800

A last even swap with monthly cost, matching the office size of the two remaining alternatives Lombard and Montana on an equal level of 950 square meters, finally enabled Alan to make his decision. After cancelling the office size objective, only two relevant objectives remained – cost and customer access. Obviously, Montana had advantages in both objectives and, dominating Lombard, could thus be revealed as the best alternative.

	Lombard	Montana
customer access within one hour (%)	80	85
office size (square feet)	~~700~~ 950	950
monthly cost ($)	~~1,700~~ 1,950	1,800

Alan signed the lease for his new office at Montana, confident that using the method of even swaps, he had made the right decision by having assessed all alternatives thoroughly and comprehensively.

Chapter 7: The generation of probabilities

7.0 Summary

1. In principle, probabilities are – at least in practical decision situations – of a subjective nature. They cannot be detected objectively but have to be elicited from people.
2. People are not used to making precise probability statements and are frequently reluctant to do so. However, in order to support the making of important decisions under uncertainty, probabilities should be quantified numerically.
3. Probabilities are elicited by direct or indirect queries.
4. An essential tool is the decomposition of complex events into basic events for which the probabilities are easier to determine.
5. With the aid of the Bayes-Theorem, prior probabilities and additional data are transformed into posterior probabilities. This is a very useful tool as humans are not intuitively capable of carrying out correct transformations.
6. A number of inadequacies of intuitive human information processing distort the process of probability generation.
7. On the one hand, these deficiencies concern the usage of incomplete or improper data.
8. On the other hand, incorrect probability judgments result from the wrong intuitive processing of probabilities.
9. A third source of error is insufficient willingness to critically view one's own probability judgments.

7.1 Interpreting probabilities

In Chapter 2, we stated how probabilities are mathematically defined: by the validity of the three Kolmogoroff axioms. However, nothing is said about how probabilities should be interpreted or measured. For many centuries, different perceptions with regard to the concept have existed (on the development of the term "probability", see Gigerenzer et al. 1990 and Hacking 2006).

7.1.1 The subjectivistic interpretation

In the subjectivistic conception, the probability of an event is the degree to which a person believes that the event will occur. It is not an objectively measurable property of the environment but rather an expression of one person's subjective conjecture about the environment.

A probability depends on the knowledge and the processing of information of an individual. This implies that different individuals may legitimately assign different probabilities to the same event. It further implies, even though this thought

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_7,

has no practical relevance, that events do not have a probability if no one forms expectations about them. Or was there a probability about "the invention of an answering machine within the next 1,000 years" at the time of Charlemagne, without anyone thinking about this?

From the subjectivistic point of view, it is irrelevant whether the uncertain event takes place in the future or has already occurred. Was Caesar ever in England? Will Turkey ever join the European Union? The first question can be answered objectively. To someone who does not know the answer, both events are uncertain and subjective probabilities may be assigned to them.

7.1.2 The frequentist interpretation

The frequentist conception of probabilities is prevalent particularly among natural scientists and statisticians. It refers to procedures which can be repeated multiple times in an identical manner. If the event e occurs in m of n repetitions of a procedure E, then e's relative frequency is given by m/n. The probability of future occurrences of the event is then deduced from the relative frequency. For example: The rate of waste in an industrial production process amounted to 2% in the last year. Thus, the probability that a component manufactured tomorrow will be defective is 2%. The probability that smokers will contract lung cancer is deduced from the relative frequencies observed in the past.

This raises the question of how many tries have to be carried out until a relative frequency can be considered a probability. In the very first final examination for business students at the Technical University of Aachen, only one candidate was admitted – and failed; the failure rate was thus 100%. Surely, this number cannot be considered as a probability for the following year's students.

Hence, some mathematicians define the probability of an event to be the limit to which the relative frequency converges in the case of infinitely many repetitions (von Mises 1951). However, since it is impossible to conduct infinitely many repetitions, a probability defined as such cannot be determined empirically; it can only be estimated from a limited number of tries.

Another problem is of great practical relevance: the frequentist interpretation of probabilities refers to identical repetitions of a procedure. Is there really such a thing? A perfect replication is barely conceivable, even if the procedures in question are highly standardized. No two spins of a roulette wheel take place under identical circumstances. If this happened and the same event occurred over and over again, roulette would no longer be a game of chance. In reality, repetitions deviate from each other to some extent. Drawing upon relative frequencies to generate probabilities for future events requires the belief that past procedures "sufficiently" resemble the future ones. Assume that, in the pharmaceutical industry, only 10% of all products have been successful in the past. Is this then also the probability of success for a new product for rheumatism? And is the rate of lung cancer in the population applicable to you, although you deviate from the statistically average person in many ways? Should you only consider a section of the overall population which resembles you in terms of age, sex, health status, smoking behavior, and so on? This raises the question of which parts of a given population one should take into account when calculating the relative frequencies.

From the perspective of a subjectivist, relative frequencies are not probabilities. However, they are always helpful as a starting point for the generation of probabilities when events are considered which have occurred repeatedly in (almost) the same manner. For example:

- Historical data on stock returns are taken into account when future returns are to be estimated.
- Knowledge of statistics on the rate of smokers contracting lung cancer may help when assessing one's own risk.
- The experience of having been caught six times when parking illegally in the vicinity of the university is relevant for the probability of being caught the eleventh time.
- With a failure rate of 60% in an exam, an only moderately prepared candidate should not be too confident of passing.

Unfortunately, in many situations in which difficult decisions have to be made, it is not possible to rely on relative frequencies because the situations are new. This is true for the following statements:

- Extending work times will improve the competitiveness of German industry.
- The Spanish galleon "Madre de Diós", which is known from old documents, lies at the bottom of the sea in grid square PQ 14.
- The controlled distribution of hard drugs to addicted persons would be a setback for organized crime.

Truly important decisions are always unique. If one were to restrict oneself to the use of frequentist probabilities in such situations, one would have to refrain from using probabilities altogether.

7.1.3 The uniform prior interpretation

Presumably all soccer fans perceive it as fair that the choice of sides is decided by a coin toss. Since the coin appears to be symmetrical, both sides of the coin are assumed to have an equal probability of landing upwards. The same applies to other simple procedures like throwing a die, playing roulette, or drawing lottery numbers. The respective mechanisms are set up in such a way that each elementary event has the same chance of occurring – or, at the very least, there is no reason to assume otherwise, unless one suspects manipulation. With n events, each of them has a probability of $1/n$. This is the classic "principle of insufficient reason" of Laplace (1812). Probabilities of more complex events can be calculated from the uniform distribution of elementary events; for example, the probability of rolling an even number with a die is $1/6 + 1/6 + 1/6 = 1/2$.

For decision situations beyond the realm of gambling, this kind of probability interpretation does not yield anything at all. For instance, if you were to flip a pin instead of a coin, it would not be obvious whether you should consider "point up"

and "point down" as equiprobable. The principle of insufficient reason would lead to a dead end if it is not possible to determine the equiprobable elementary events.

7.1.4 Subjective and objective probabilities

In contrast to the concept of subjective probabilities, which prevails in modern decision theory, many authors consider frequentist and symmetric probabilities to be "objective"; objectivity is a desirable property. Since scientists seek objectivity – that is, intersubjective verifiability – an aversion to the subjectivistic concept of probability is understandable. We have outlined the subjectivistic perspective already; are there objective probabilities as well?

How strict the view on objective and subjective probabilities can deviate from one another is demonstrated by both of the following quotations.

> The probability subjectivism is of course absolutely incompatible with the attempt to gain insights into objective facts and interdependencies. [...] This interpretation of probabilities by the individual – for whose utility calculus they are interesting – touches the probabilities as little as the paths of the stars are touched by the fact that we set our watches according to them. Talking about subjective probability is as meaningful as talking about subjective causality. (Menges 1968, p. 34).

> Yet, despite the very wide acceptance of the frequentist view, I and many others consider that in all circumstances it is philosophically unsound or, to be more blunt, arrant nonsense. [...] The subjective view of probability, on the other hand, does fill our needs in decision analysis and, I shall contend, is philosophically sound. (French 1988, p. 234).

From our perspective, an objectivity of probabilities is unobtainable in a stricter sense. Relative frequencies may be objective, but they are not probabilities. To extrapolate from known frequencies to unknown cases is a subjective hypothesis; it cannot be ascertained objectively whether this hypothesis is appropriate. However, integrating objective circumstances may simplify the process of quantifying probabilities The larger the extent to which probabilities are based on objective empirical data, the more willing different decision makers will be to agree on these probabilities. If one is to acknowledge the agreement of many people as a sign of objectivity – which is problematic enough – then many probabilities gain a higher, although less than 100%, degree of objectivity.

If at all, the uniform prior interpretation building on simple random mechanisms may be regarded as "objective". If one of a hundred table-tennis balls is to be drawn out of a container after carefully mixing the balls, all observers will agree that each ball is drawn with a probability of 1/100, as long as they do not suspect manipulation. However, in the history of mankind, many a die has been manipulated; many decks of cards have been marked. At the beginning of 1999, manipulation in the Italian lottery emerged - the blindfolds of the orphans who drew the balls were not opaque, and to play it safe, some balls were heated as well to make them feel different to the others. The unsuspecting lottery gamblers certainly had no reason to assume different probabilities for the lottery numbers. However, their hypothesis was purely subjective and ex-post proved to be incorrect.

There is no definition to be found for objective probabilities; indeed, they are a contradiction in terms. How "true" something "appears" to be cannot be determined objectively per se. If at all, it is possible only to determine whether something *is* true or not. If this is known, then there is no uncertainty and there will be no room for probabilities (apart from 1 or 0).

Regardless of whether or not one follows this view and abandons the concept of objective probabilities altogether, for practical reasons one has to use subjective probabilities as in real-life decision situations, frequentist or uniform prior concepts do not prove to be helpful.

Subjectivity does not mean arbitrariness. Subjective probabilities have to satisfy the Kolmogoroff axioms listed in Section 2.3.1. Also, the procedures used to determine the probabilities have to be based on axioms. An axiomatic reasoning for the measurement of subjective probabilities was given by Savage (1954, 1972). Other statements can be found in Raiffa (1968), Lindley (1971, 1991) and French (1988).

Apart from this consistency requirement, requirements have to be set for the factual quality of subjective probabilities. We have mentioned this fact in Chapter 1, using the term "procedural rationality". Probabilities should be *founded*; they should rest on an informational base which is appropriate to the problem. The decision maker or the expert who states a probability should be able to explain the reasons that lead him to his estimate in a way that is comprehensible to others.

There are numerous methods which may contribute to a better modeling of uncertainty (Russo and Schoemaker 1989, Kleindorfer et al. 1993):

- Listing all pros and cons for a certain probabilistic statement,
- Generating causality trees to recognize possible sources of future failure,
- Creating optimistic and pessimistic scenarios which may result from the concurrence of multiple influencing factors,
- Anticipation of a future setback and its pseudo-retrospective explanation (*prospective hindsight)*.

7.2 The need to quantify probabilities

Unconsciously, we continuously form expectations about events and use those expectations in our decision making. If you speed through a red light, you probably perceived the probability of a crash as insignificantly low. If you are just about to leave the house and the telephone rings, you will consider the probability of the call being important and then decide whether to answer the phone. More serious decisions require a more careful evaluation of opportunities and risks.

Expectations about uncertain events are mostly generated and communicated in a vague verbal way, as follows:

- A trade embargo between the US and China is very unlikely at the moment.
- It seems like the US-Dollar will fall further in the next few days.
- It is not impossible that the electromagnetic radiation from mobile phones is harmful.
- It cannot be assumed that *Schalke* will win the football league this season.

Such verbal expressions are not only imprecise, but even more problematically are understood differently from person to person. In a survey, 40 English-speaking managers were asked to rank ten expressions in descending order of certainty. On average, "quite certain" was named the term with the highest certainty and "unlikely" the term with the highest degree of uncertainty. However, the wide range of classifications was remarkable; for example, "likely" was named at positions from the second to the seventh, "probable", from the second to the ninth, "possible" from the third to the ninth and "unlikely" from the third to the tenth place. The authors of this present book obtained similarly striking results when conducting comparable evaluations with the students in their lectures about decision theory. This can be seen in Figure 7-1 which shows the spread, the 10%-quantile and the 90%-quantile, as well as the median of evaluation of the different expressions by the 41 respondents. For example, the expression "possible" (in German: möglich) was associated with probabilities of 1% to 80% by different individuals. The 10%-decile was 20%, the 90%-decile 60% and the median 50%. "Unlikely" (unwahrscheinlich) varied between 1% and 20%, whereas the 10%-decile constituted 5% and the 90%-decile 20%. For the statement "to be expected" (zu erwarten), probabilities between 60% and 98% were stated; the corresponding deciles were 73% and 93%, respectively, with a median of 80%.

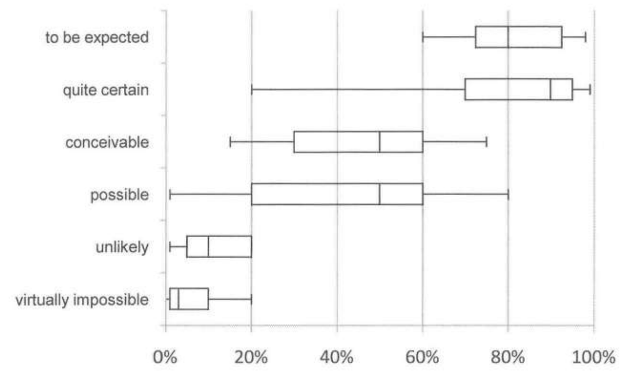


Figure 7-1: Numerical interpretations of different probability terms

In everyday life, achieving a high degree of accuracy is rarely crucial; it does not pay off to invest in precision. However, quantitative probability statements are recommended for important decisions. They have the advantage of being intersubjectively unambiguously understandable; they can be compared to each other and discussed in a sensible manner. One can use them for calculations as their informational content can be processed easily.

Surprisingly, it is not only in colloquial language where quantitative probability statements are found less frequently than vague verbal probability statements. Particularly in the professional area, where expectations about uncertain events are of critical relevance, there seems to be a reluctance to quantify such expectations as a matter of principle. This is the case even though one would expect professional decision makers to know about the concept of probabilities. Abundant examples can be found in the fields of medicine or law.

In package inserts, one can read horrifying lists of possible adverse reactions of seemingly harmless pharmaceuticals. The probabilities of the occurrence of such reactions is expressed only in vague expressions like "a small percentage of patients" or "infrequently". The decision as to whether to take the medicine is not supported rationally by such statements. Some patients might refuse to take the drugs with the possible adverse reactions, as they are not able to trade the advantages off against the risks due to their inadequate knowledge of the relevant probabilities. In an experiment by Fischer and Jungermann (1996), subjects associated the term "rare" in the context of adverse reactions to a drug with a probability of 2.3%, "uncommon" with 6.8%, and "common" with 23.8%. The same terms, outside of a medical - or indeed any specific context - were on average interpreted as follows: "rare" 12.8%, "uncommon" 27.8%, and "common" 66.8%. This shows that the perception of vague probability statements depends heavily on the respective context. In the meantime, the need to quantify probabilities has been acknowledged, at least in principle. In its „Recommendation for the design of package inserts" (Bekanntmachung von Empfehlungen zur Gestaltung von Packungsbeilagen, November 2006) the Federal Institute for Drugs and Medical Devices states:

The frequency of adverse drug reactions according to European recommendations:

Very common:	more than 1 of 10 treated persons
Common:	less than 1 of 10, but more than 1 of 100 treated persons
Uncommon:	less than 1 of 100, but more than 1 of 1000 treated persons
Rare:	less than 1 of 1000, but more than 1 of 10000 treated persons
Very rare:	less than 1 of 10.000 treated persons
Unknown:	Frequency cannot be estimated due to insufficient data

By now, many package inserts contain this additional information which clarifies how the terms "very rare" to "very common" are to be interpreted in this context.

Probability estimates concerning existential questions can be found, for example, in the context of an advance health care directive. In this case, it is difficult for the individual to predict in which situation she would refrain from certain medical treatments or forecast the probabilities of their success. The German Ministry of Justice has published a series of text modules which describe possible situations

in which advance directives may be exercised (Bundesministerium der Justiz 2010):

(a) in all likelihood I am in the irreversible, imminent process of dying...
(b) I am in the terminal stage of an untreatable terminal disease, even if the time of death is not yet foreseeable...
(c) due to a brain injury, my ability to gain insights, make decisions and make contact with other people is, in the opinion of two experienced physicians, in all likelihood irreversibly lost, even if the time of death is not yet foreseeable [...] I am aware that, in such situations, the ability to feel may still be intact and that recovering from this condition may not be ruled out with complete certainty, but is improbable...

Here, terms such as "in all likelihood" and "is improbable" are used; of course, an exact probability forecast is intrinsically complex. However, the patient possibly stakes her life on how physicians - not personally known to her - interpret these terms. In addition, it might only be possible for the doctors to make the relevant estimates imprecisely. Other organizations have developed their own text modules with comparable criteria which are just as difficult to interpret; this makes a uniform treatment of such terms even more difficult (for an overview see Jacobi et al. 2005).

Jurists have notorious difficulties with the quantification of expectations; *„Iudex non calculat"*. Judges frequently have to render judgments under uncertainty. In doing so, the law does not provide them with much help. Scholl (1992) gives a couple of examples; one of them is § 252 of the German civil code. The law reads as follows:

The damage to be compensated for also comprises the profit lost. the profit considered lost is the profit that could probably have been expected in the normal course of events, or in the special circumstances, particularly due to the measures and precautions taken..

This phrasing leaves it to the jurisprudence to clarify what "probable" means and does not bother about the fact that different profit magnitudes are usually to be expected with different probabilities.

Scholl remarks that courts are reluctant to using fixed probability values as criteria. In a case of vaccination damage, the mandatory immunization against typhoid was ruled to be the cause of a sarcoma although the probability that a sarcoma follows from a typhoid immunization was less than 0.01%. The judges argued that the question of adequacy between cause and effect could not be answered on a purely logical and abstract basis according to the frequency of occurrence of such events. Instead, drawing on judgmental considerations, a limit had to be found up to which one can justly expect the originator of the cause to bear the liability for its consequences (Scholl 1992).

Vague expressions not only conform to the intuitive manner in which we sense probabilities; they can also serve as smoke screen - an underlying lack of complex thinking and a lack of willingness to assume responsibility can be disguised in this manner. Furthermore, arguing for or against alternatives in a vague manner without revealing one's own preference is facilitated: assume that persons X and Y are applying for a job. X is less qualified, but a good friend of yours. You throw into the debate that it is to be expected that X will soon become acquainted with the complex matters and thus compensate for his deficits. A statement of this manner is generally understood as non-binding, you can use it without any risk. But would you be willing to commit to a probability of 95% if you were not *really* very sure?

If we insist that probabilities have to be quantified in difficult decision situations, then this does not mean that we insist on point estimates. It is often difficult to state an exact value for the probability of some event, but giving intervals for a probability may also prove very helpful as such intervals are frequently sufficient to identify the best alternative. To us, a doctor's statement "The risk of dying during this surgery is between 1 and 3%" seems more useful than "Do not worry, in most cases, there are no problems." In chapter 10, we will cover the treatment of incomplete probability information, i.e. probabilities that are not given as point estimates.

7.3 The measurement of subjective probabilities

7.3.1 Probability and distribution functions

Since probabilities are merely mental constructs, they need to be elicited from people's minds. Unfortunately, it is generally not possible to simply ask people about numerical probabilities. Most individuals are not accustomed to dealing with probabilities: first and foremost, at the beginning of the process, the desired probabilities are not yet existent or only vaguely known; the person in question has not sufficiently contemplated them. However, questioning methods have been developed which help individuals (decision makers or experts) to express and specify their notions of a specific probability. (Spetzler and Staël von Holstein 1975, Morgan and Henrion 1990, Meyer and Booker 2001).

The decision analysis literature assumes that interrogator and respondent are different persons. The interrogator should understand probability calculus, the methods of probability measurement, and the usual biases which may be observed. (see 7.5). The respondent should be the expert, meaning that he or she should possess knowledge on the practical problem at hand. Of course, it cannot be ruled out that, both requirements may be fulfilled by one single person in specific cases.

On the one hand, probability measurements may concern *single events*, conditions, or scenarios. This could for instance be the probability that our sponsored tennis player joins the world's top ten in the next year, or that an earthquake with the magnitude of 7 on the Richter scale will devastate the San Francisco Bay Area in the next 30 years. On the other hand, probability distributions of numerical variables are often considered. These are termed *random variables.* For example, one might be looking for the probability distribution of the sales volume of a product, of some US-Dollar exchange rate or of the oil price.

Distributions of random variables may be described either by probability (also known as density) functions or by distribution functions.

Every possible value x_i of a *discrete* random variable has a positive probability p_i. The function $p(x)$ which allocates a probability to each number x is called the probability function;

$$\begin{aligned} p(x) &= p_i,\ \text{if } x = x_i \\ &= 0 \text{ otherwise.} \end{aligned}$$

The distribution function $P(x)$ specifies the probability that the random variable will take on a value smaller or equal to x. It holds that:

$$P(x) = \sum_{x_i \le x} p_i \tag{7.1}$$

Assume, for example, that one wants to estimate the distribution of the number of operations per night of a certain fire department. Probabilities are assigned to the possible values 0, 1, 2, ... , x_{max}. A suitable graphical illustration is given by a bar diagram (Figure 7-2a).

In the case of *continuous* random variables, there are no positive probabilities for singular values, only for intervals. Accordingly, one can say that the water level of the Rhine in Cologne tomorrow will be between 8.5 and 8.6m with a probability of 90%. Instead of a probability function, they are defined via a distribution function; the probability of an interval is equal to the integral of the distribution function over that specific interval. It often proves useful to identify intervals for which the probability density is constant and to assign probabilities to them; the density function can then be displayed by a bar chart (histogram); the area of each bar is proportional to the probability that the value of the variable falls into the interval (Figure 7-2b).

The simplest distribution of this kind is a uniform one which is characterized by a, density that is constant over the entire area between the upper and the lower limit (Figure 7-2c).

A simple distribution with a constantly increasing and then constantly decreasing density is called a triangular distribution (Figure 7-2d).

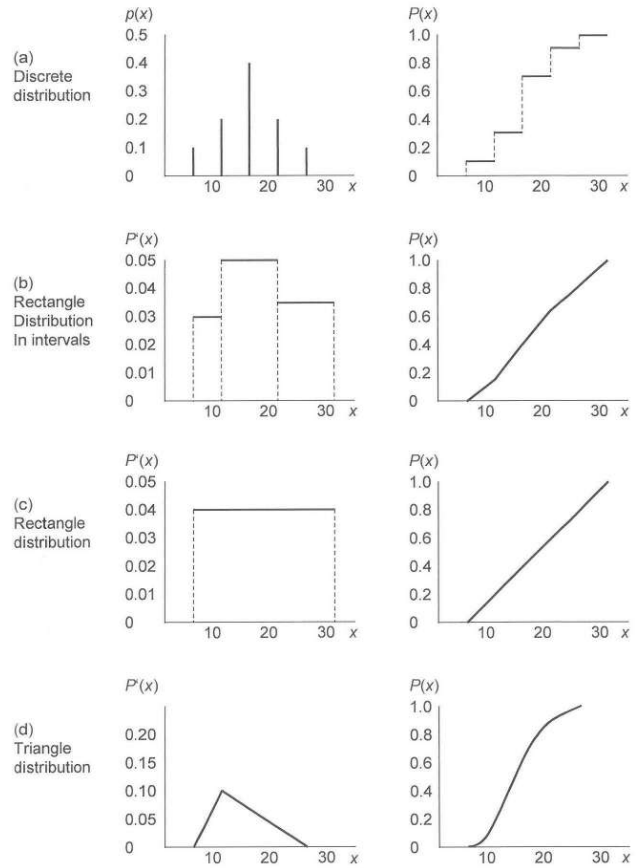


Figure 7-2: Some probability or density functions (left) with the associated distribution functions (right)

We pointed out in Chapter 2 that a variable is discrete or continuous "by nature". However, its representation in the decision calculus may be different. It is often necessary to choose a discrete representation for a continuous variable for the purposes of simplification. In addition, a discrete or a continuous approach might be chosen when *measuring* the distribution, independently of the variable's "true na-

ture" or its subsequent representation. For example, let's examine the uncertain variable "number of newly registered automobiles in May". This variable is inherently discrete, but because of the thousands of possible values, one could measure this variable as if it were continuous and construct a density like in Figure 7-2, (b) through (d). The continuous distribution function can then be discretized again for the decision calculus, e.g. reduced to three or five quantiles so that the problem can be represented in a decision tree or matrix.

The functions shown in Figure 7-2 are "all-purpose distributions" which often adequately model experts' probability perceptions of various situations.

In addition, there exists a wide variety of other theoretical distributions which may be appropriate for some special cases. If experts have reason to assume a particular type of distribution, such as a normal or binomial distribution, then determining the specific distribution requires only estimation of the respective parameters. If possible, one will derive the estimates from a sample. However, in novel situations, one is forced to use subjective forecasts.

7.3.2 Measurement methods

A distinction needs to be made between direct and indirect measurement methods. With direct methods, the respondent answers the questions of the interviewer by providing a figure, that is, by specifying a probability or the level of the uncertain variable for a given probability y. With indirect methods, the respondent draws comparisons with simpler situation so that the probabilities may be derived in this manner. This concept relies on the assumption that the respondent is not familiar with the idea of probabilities and therefore needs help in order to make meaningful appraisals.

Furthermore, measurement methods can be distinguished depending on whether probabilities or levels of the uncertain variable are asked for. There are hence four possibilities, which are listed in Table 7-1 and will be outlined below.

Table 7-1: Alternative methods of measuring probabilities

	Direct measurement	Indirect measurement
Query for probabilities	Direct query for probabilities	Indirect query for probabilities
Query for values of the uncertain variable	Direct query for values	Indirect query for values

Direct query for probabilities

For discrete events, the interviewer asks for their probabilities straight away, for example: "What is the probability that a night frost will occur before the harvest takes place?", or "What are the probabilities of zero, one, two… operations of the fire brigade next weekend?" It may prove helpful to let the respondent order the events according to their probability before quantifying them. Additionally, the respondent may visualize the proportion of the probabilities using graphs (bar or pie chart).

A distribution function can be determined for continuous variables. If the quantity demanded is to be estimated, one could provide the respondent with the intervals

- 10,000 bis 11,000 metric tons,
- 11,000 bis 12,000 metric tons,
- 12,000 bis 13,000 metric tons,
- 13,000 bis 14,000 metric tons,
- 14,000 bis 15,000 metric tons

and ask him to allocate probabilities to these intervals. Afterwards, one could ask him if he assumed a uniform distribution within the intervals. If this is not the case for some individual intervals, these intervals will be split up. If the respondent assigns a probability of 0.5 to the area between 12,000 and 13,000 metric tons, then he will be asked for the probability of an amount between 12,000 and 12,500 tons. Eventually, one obtains a set of intervals for which a uniform distribution can be assumed.

It is important to ensure that the intervals cover the complete range of possible levels of the respective variable. As a precaution, one should initially define a very broad interval.

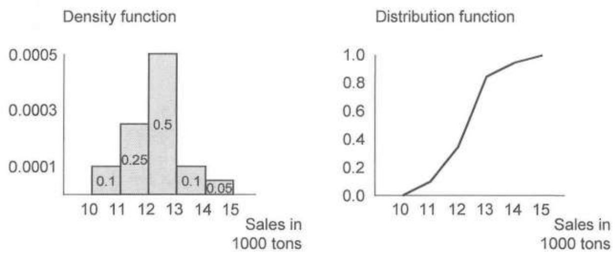


Figure 7-3: Density and distribution functions

If the i-th interval is given by $x_{i0} \leq x \leq x_{i1}$ and assigned a probability of p_i, then $p_i = (x_{i1} - x_{i0}) \cdot P_i'(x)$ and it holds that the interval has a constant density of $P_i'(x) = p_i / (x_{i1} - x_{i0})$. For example, a sales volume of 11,000 to 12,000 tons has a probability of 0.25. The interval has the density 0.25/(12.000 – 11.000) = 0,00025. In Figure 7-3, a possible result of the query is depicted; the shaded areas beneath the density function represent the probabilities of the intervals.

The distribution function can be derived from the density. The numbers in the example result in the distribution function in Figure 7-3.

Instead of using the density, the distribution function can be determined directly. For this purpose, one may ask questions of the following kind: "What is the

probability that our sales will be less than or equal to €200 million next year?" The answer yields one point on the distribution function as plotted in Figure 7-4. With this method, one should be careful in evaluating the upper and lower limits. If the respondent states a probability of zero when asked for the probability that the sales fall below 50, and he further states a probability of 20% that the sales fall short of 100 then the lower limit will lie somewhere between 50 and 100. In order to more exactly localize the limit, the interval has to be split and the query needs to be extended to cover these new subintervals.

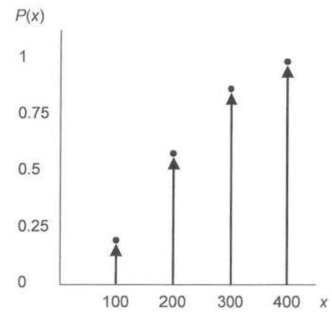


Figure 7-4: Query for the probabilities of a distribution function

Using this method yields a limited number of points of the distribution function; these points can then be used in a variety of ways to generate a complete function. The points may be connected linearly, or be approximated by a free-hand line or a theoretical curve, or one might simply assume a step function. Figure 7-5 shows an approximation by a smooth function. This can now be used to read off the probability that a random value will fall within a certain interval. In Figure 7-5, the probability for $x_1 \leq x \leq x_2$ is equal to $P(x_2) - P(x_1)$. If you wanted to discretize the distribution, it would make sense to allocate this probability to the x-value which would halve the probability of the interval.

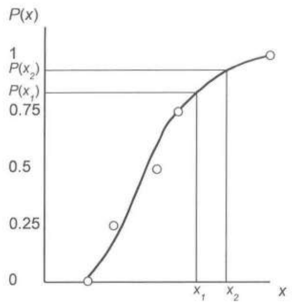


Figure 7-5: Approximation of elicited points by a smooth distribution function

Direct queries for values

The method of direct queries for values also attempts to elicit specific point values of the distribution function. In this method, however, probabilities are provided and corresponding levels of the uncertain variable are queried for. Accordingly, questions of the following kind are posed: "What sales volume will the actual sales exceed and fall short of with equal probability?" Or "What sales volume will be exceeded with a probability of 10%?" The answers to both questions yield the median and the 90%-quantile of the distribution.

Experience shows that it is advisable not to start with the median but with the extreme levels instead. If the median is considered at the beginning, people will tend to underestimate the bandwidth of the variable; the distribution hence becomes too narrow (see 7.5.4 below). One may proceed as follows:

1. Evaluation of the upper and lower limits. ("What is the minimal sales volume, what is the highest possible?") or of the 5% and 95%-point.
2. Evaluation of the median. ("What sales volume will the actual sales exceed and fall short of with equal probability?")
3. Evaluation of the lower quartile. ("For which value is it equally probable that the sales will fall below this value or that they fall between this value and the median?" Or: "For which value is the probability that the sales will exceed this value three times as high as the probability that the sales will fall short of this value?"
4. Evaluation of the upper quartile (analogous to 3).

This procedure corresponds to the bisection-method for the evaluation of value functions as outlined in Chapter 5. Figure 7-6 gives a schematic depiction.

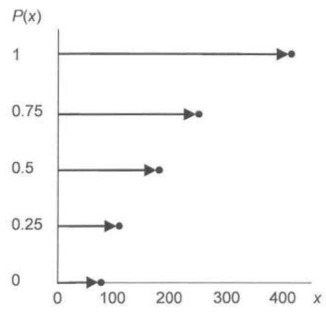


Figure 7-6: Query of quantiles of a distribution function

This method is appropriate for continuous and discrete variables with many possible values but not for discrete variables with relatively few values. This is because it is unclear if the respective quantiles exist. For example, the number of fire alarms in a city is a discrete variable. Assume that an expert estimates a probability of 40% for the value "no alarm" and 30% for "one alarm". Neither the median nor the 25%-point exists.

Indirect query for probabilities

Most people are not used to handling numerical probabilities and have difficulties stating them. The indirect methods avoid forcing the respondent to explicitly state probabilities or precisely understand their meaning.

For example, assume that one is interested in the probability estimated by an engineer that the prototype of the new model will be finished by the end of 2010. If the expert is not familiar with the quantification of probabilities, one may stimulate his imagination by equipping him with a bucket containing nine white ping pong balls and one yellow ball and give him two lotteries to choose from: You can win €1,000 if the prototype is finished, or €1,000 if at the end of the year, the yellow ball is drawn. Which bet would you prefer? If the expert decides to choose the bet on the construction of the prototype in time, he implicitly estimates this event's probability to be greater than 0.1. The interviewer will now replace one white ball with a yellow ball so that the bucket now contains two yellow ping pong balls and eight white ones, and then ask the expert the same question again. This is repeated until the respondent is indifferent between both bets. Assume that he feels that the bets are on par for four yellow and six white balls; this means that his estimated probability of the prototype being finished in time amounts to 0.4. Figure 7-7 illustrates this case.

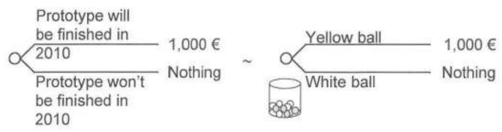


Figure 7-7: Indirect measurement of the probability of an event

The advantage is that the term "probability" does not have to be used at all. In addition, the respondent might make a greater effort if some money is at stake – even though it is just virtual.

In place of a bucket with ping pong balls, the literature often recommends a probability wheel; this is a disc consisting of two differently colored segments and a static pointer or hand (Figure 7-8). The respondent is asked to imagine that the disc were to be turned; when it comes to rest, the hand will either be in the blue or in the yellow sector. The relative size of the two colored sectors represents their probability and the size of the winning yellow sector is changed until the respondent is indifferent between a bet on the yellow sector and one on the event of interest – e.g. constructing a prototype before the end of 2010.

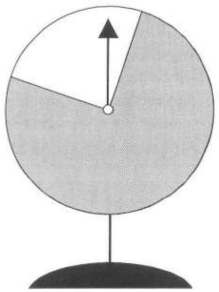


Figure 7-8: Probability wheel

Regarding practical application, the probability wheel has the advantage over ping pong balls or deck of cards that the probabilities can be varied continuously. A disadvantage (apart from the fact that it is difficult to find one in a shop) stems from the fact that very small probabilities cannot be represented adequately with the wheel. A probability of 1/10,000 may better be visualized by the boxes of a square of 100 x 100mm on millimeter paper.

The example of the prototype concerned the measurement of probabilities of discrete events. In the same way, density and distribution functions may be measured indirectly. The probability that the uncertain variable will fall into a given interval is equal to p if an expert is indifferent between a bet on this event or another

bet with the same prize and a given probability of p; one can accordingly determine the cumulative probabilities $P(x)$.

Indirect query for values

Here, a reference situation with a fixed probability is used, such as a coin toss or a draw of a ball from a bucket containing one yellow and nine white balls. The level of the distribution that is to be elicited is varied. For instance, to find the 25%-point of the distribution of sales, you let the marketing expert choose whether he rather wants to bet that the next card to be drawn is diamonds (with a probability of 25%) or that the sales volume will fall short of, for instance, €100,000. The prize is the same for both cases. If the expert prefers to bet on diamonds, this indicates that the probability of falling short of €100,000 in sales is less than 25%. Accordingly, you then increase the sales volume, for example up to €150,000. Assume that the expert now prefers to bet that the actual sales will fall short of this amount; this implies that the probability of falling short of €150,000 has to be greater than 25%. The sales volume which makes the expert indifferent between the two lotteries is thus the 25%-point.

You can find the median by using an analogous procedure. You let the expert decide if he would rather bet on a coin toss landing on "tails" or that the sales volume will fall short of some amount, let's say €200,000. If the expert prefers to bet on the sales missing the sales target, his subjective probability for the event is greater than the probability of "tails", i.e. 0.5; in other words, the median is below €200,000 and we hence have to lower the value to reach indifference. Assume the expert is indifferent between betting on the coin toss and betting on the sales volume falling short of €175,000; this sum is then the median of the distribution.

To determine the 75%-point, it would be necessary to create a reference lottery which would yield a probability of winning of 75% - for instance drawing clubs, spades or hearts from a deck of cards.

7.3.3 Consistency checks and the reduction of error

The process of measuring probabilities is prone to error, making consistency checks important. All probability judgments made by one single person should be consistent with each other, and the points of a distribution function should not deviate too much from a monotonically increasing function.

Consistency checks may be conducted within one and the same method or by querying for probabilities using multiple methods. Therefore, it is advisable to query both for values and for probabilities. If the points yielded by the two methods do not match, a revisal becomes necessary. In addition to the probabilities of certain events, the complementary probabilities may be asked for; the sum must be equal to one. With complex combinations of events, decomposition into simpler events will frequently prove helpful. For the composition, the rules of probability calculus can then be used. A well-known example was already mentioned in Chapter 2, the estimation of accident risk in nuclear reactors in the so-called Rasmussen report.

In Section 7.5, we will describe a series of systematic errors which can be observed repeatedly in connection with probability judgments. Not all of these biases can be eliminated by consistency checks. An awareness of the widespread existence of systematic biases is a prerequisite for reducing their influence. The danger of bias will in general be mitigated if interviewer and respondent are different persons. A "self-interrogation" may suffice for some less important personal decisions; beyond that, the process will be less prone to error if the elicitation is conducted by an objective person equipped with the necessary methodological knowledge.

7.3.4 The calculation of probabilities

Decomposition is a basic principle of prescriptive decision analysis, which also holds true for the determination of probabilities. The probability of a complex event will often be quantified more easily if it is perceived as a combination of other events; the probability of a complex event is then calculated by the rules of probability calculus. We have shown this already in Chapter 2, using the example of the risk estimation for nuclear reactors. We have also already pointed out the use of event trees and fault trees.

The Bayes-theorem is an important special case of calculating probabilities by the conjunction of events. In the following section, we will elaborate on this concept.

7.4 Bayes' theorem

Bayes' theorem (1763) is a formula that allows the computation of probabilities from other probabilities. Its specific strength lies in its ability to revise initially assumed ("prior") probabilities in the light of new data.

Let us consider uncertain environmental conditions s_1, s_2, ..., s_n, to which we have assigned the prior probabilities $p(s_1)$, $p(s_2)$, ..., $p(s_n)$. Furthermore, there is one single source of information at our disposal. This source provides us with exactly one piece of information (in the following analysis also referred to as an observation) from the information set $Y=\{y_1, y_2, ..., y_m\}$. We know the conditional probabilities $p(y_j|s_i)$ that y_j is observed given that s_i is the true condition. The case is shown in Figure 7-9.

We obtain the joint probabilities $p(y_j,s_i)$ by

$$p(y_j, s_i) = p(s_i) \cdot p(y_j|s_i). \tag{7.2}$$

We now observe y_j. How can y_j be used to revise the prior probabilities of the environmental conditions?

Following the observation ("*posterior*"), the different environmental conditions generally have different probabilities. They have now become the conditional probabilities

$$p(s_i \mid y_j) = p(y_j, s_i) / p(y_j). \tag{7.3}$$

The right side of 7.2 will go into the numerator; the marginal probability of observation y_j into the denominator. The marginal probability consists of the sum of the joint probabilities of all event combinations in which y_j occurs so that it adds up to

$$p(y_j) = \sum_i p(y_j, s_i)$$

This gives us

$$p(s_i \mid y_j) = \frac{p(s_i) \cdot p(y_j \mid s_i)}{p(y_j)} \tag{7.4}$$

or

$$p(s_i \mid y_j) = \frac{p(s_i) \cdot p(y_j \mid s_i)}{\sum_i p(s_i) \cdot p(y_j \mid s_i)}. \tag{7.5}$$

This is Bayes' theorem. It can be used to transform the prior probabilities for the uncertain environmental conditions s_i into posterior probabilities with the aid of conditional probabilities for the observation y_j. The conditional probabilities $p(y_j|s_i)$ are called *likelihoods*.

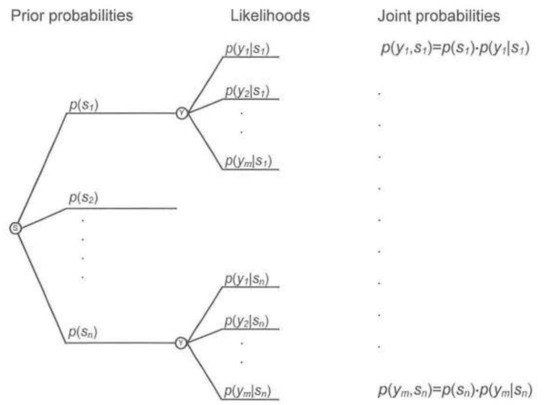


Figure 7-9: Prior probabilities and likelihoods

Use the following simple case, also illustrated in Figure 7-10, to begin understanding the relationships: A man has been arrested under suspicion of murder. Evidence collected so far suggests a prior probability of 0.7 that he is guilty. A polygraph ("lie detector") test is used in the hope that it will produce further information. However, polygraphs are, not infallible. They correctly identify liars in 90% of the cases, correctly showing a positive result; in 10% of the cases, they falsely show a negative result. Conversely, in 20% of the cases, they falsely accuse an innocent person and correctly show a negative result in 80% of cases. We will assume these (fictitious) empirical frequencies as probabilities for this case.

The suspect undergoes the lie detector test. Suppose that he passes, i.e. is tested negative. What is now the posterior probability that he is guilty? We have prior probabilities:

$$p(\text{guilty})=0.7 \quad p(\text{innocent})=0.3$$

and the likelihoods

$$p(\text{test pos.}\,|\,\text{guilty})=0.9 \quad p(\text{test pos.}\,|\text{innocent})=0.2$$
$$p(\text{test neg.}\,|\,\text{guilty})=0.1 \quad p(\text{test neg.}\,|\,\text{innocent})=0.8.$$

This yields the joint probability

$$p(\text{guilty, test neg.})=p(\text{guilty})\cdot p(\text{test neg.}\,|\,\text{guilty})=0.7\cdot 0.1=0.07$$

and the (unconditional) probability that the test will be negative

$$\begin{aligned}&p(\text{test neg.})=\\ &p(\text{guilty})\cdot p(\text{test neg.}\,|\,\text{guilty})+p(\text{innocent})\cdot p(\text{test neg.}\,|\,\text{innocent})\\ &=0.7\cdot 0.1+0.3\cdot 0.8=0.31.\end{aligned}$$

Therefore, it holds for the posterior probability that

$$\begin{aligned}p(\text{guilty}\,|\,\text{test neg.})&=\frac{p(\text{guilty, test neg.})}{p(\text{test neg.})}=\frac{0.07}{0.31}\\ &=0.226.\end{aligned}$$

Would the suspect fail to pass the test (i.e. test positively), the probability him being guilty will increase to

$$\begin{aligned}p(\text{guilty}\,|\,\text{test pos.})&=\frac{0.7\cdot 0.9}{0.7\cdot 0.9+0.3\cdot 0.2}\\ &=0.913.\end{aligned}$$

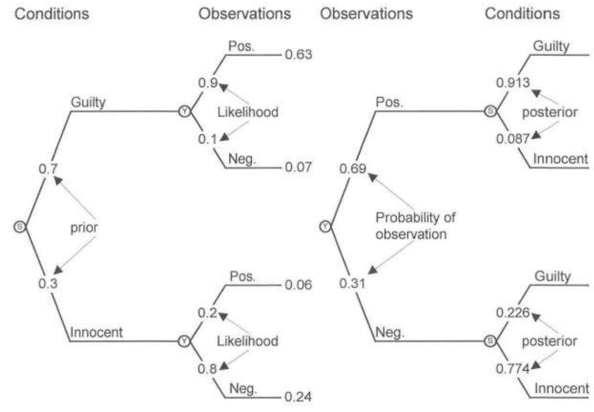


Figure 7-10: The murder suspect and the lie detector

The two probability trees in Figure 7-10 show the situation from different angles. Tree number one shows the determination of the joint probabilities of conditions and observation data. Tree number two displays the data probabilities and the posterior probabilities of the conditions.

Obtaining likelihoods

The conditional probabilities of specific information can (like all probabilities) be obtained in different ways, more specifically

- on the basis of empirical relative frequencies for recurring events. For example, patient data of indicators and symptoms are collected in medical histories. The diagnosis for newly admitted patients will be guided by the probabilities derived from their indicators and symptoms.
- on the basis of theoretical distributions assumed to be suitable. For example, let r be the unknown scrap rate in a production process. The probability that y of n randomly picked work pieces are faulty is distributed binomially:

$$p(y\,|\,n,r)=\binom{n}{y}\cdot r^{y}\cdot(1-r)^{n-y}. \tag{7.6}$$

  The prior distribution of the scrap rate can thus be revised by the help of a random sampling result y.
- through intuition. For example, by finding an answer to the question: what is the probability that the doctor who has examined me for cancer will tell me the truth about whether this condition is fatal?

Computation table

For a quick and simple computation of Bayes' theorem, the table shown in 7.2 (Samson 1988, p. 352) has proven particularly useful. You first enter the prior probabilities $p(s_i)$ of the uncertain conditions and the likelihoods of the data for given conditions in section 1 of the table. According to the formula $p(y_j,s_i)=p(s_i)\cdot p(y_j|s_i)$, you then compute the joint probabilities and enter them in section 2. By summation over each row, you obtain the prior marginal distribution of the conditions in the right table margin; accordingly, you obtain the marginal distribution $p(y_j)$ of the data in the bottom table margin.

The posterior probabilities can now be calculated by dividing the joint probabilities $p(y_j,s_i)$ for each observation y_j by the unconditional probability $p(y_j)$; enter these numbers in section 3of the table.

For clarification, you will find a sample calculation underneath. For a single-handed bank robbery, the usual suspects Alfred (A) and Bertie (B) are considered as possible perpetrators. The police estimate a 30% probability that Alfred is guilty, a 10% probability for Bertie and 60% for another bank robber (*?*).

An eyewitness saw the man escape after the robbery with his mask removed. The witness is confronted with Alfred and Bertie in a police line-up. His possible statements are: *a* (it was Alfred), *b* (it was Bertie), or *?* (don't know).

Since Alfred's looks are very distinctive, the police reckon there is a probability of 80% he will be recognized by the witness if he is indeed the robber. A probability of 10% each is assigned to statements *b* and *?*. The other likelihoods must be interpreted analogously. Bertie looks rather ordinary so the probability of his being recognized is estimated at only 50%; a witness who has seen Bertie will identify Alfred with a probability of 10%, and with a 40% probability he will recognize neither of the two. Should neither of the two actually be the robber, there is a 15% probability of the erroneous identification of Alfred, and a 30% probability of the erroneous identification of Bertie.

The calculation shows: if the witness incriminates Alfred, Alfred's suspiciousness increases from 0.3 to 0.71 while Bertie's decreases from 10% to 3% and the probability of a third man is at 26%. Should the witness, however, believe he recognizes Bertie, the probability of him being the offender only increases from 10% to 19%. The probability that Alfred is the robber is 12% while that of a third party is 69%.

Table 7-2: Table for the computation of posterior probabilities

		1						2						3				
		Data						Data						Data				
		y_1	…	y_j	…	y_m		y_1	…	y_j	…	y_m		y_1	…	y_j	…	y_m
Conditions	prior	Likelihoods					Σ	Joint probabilities					Σ	Posterior probabilities				
s_1	$p(s_1)$	$p(y_1\|s_1)$	…	$p(y_j\|s_1)$	…	$p(y_m\|s_1)$	1	$p(s_1,y_1)$	…	$p(s_1,y_j)$	…	$p(s_1,y_m)$	$p(s_1)$	$p(s_1\|y_1)$	…	$p(s_1\|y_j)$	…	$p(s_1\|y_m)$
⋮	⋮	⋮		⋮		⋮		⋮		⋮		⋮	⋮	⋮		⋮		⋮
s_i	$p(s_i)$	$p(y_1\|s_i)$	…	$p(y_j\|s_i)$	…	$p(y_m\|s_i)$	1	$p(s_i,y_1)$	…	$p(s_i,y_j)$	…	$p(s_i,y_m)$	$p(s_i)$	$p(s_i\|y_1)$	…	$p(s_i\|y_j)$	…	$p(s_i\|y_m)$
⋮	⋮	⋮		⋮		⋮		⋮		⋮		⋮	⋮	⋮		⋮		⋮
s_n	$p(s_n)$	$p(y_1\|s_n)$	…	$p(y_j\|s_n)$	…	$p(y_m\|s_n)$	1	$p(s_n,y_1)$	…	$p(s_n,y_j)$	…	$p(s_n,y_m)$	$p(s_n)$	$p(s_n\|y_1)$	…	$p(s_n\|y_j)$	…	$p(s_n\|y_m)$
							Σ	$p(y_1)$		$p(y_j)$		$p(y_m)$	1	1		1		1

		1				2				3		
		Witness account				Witness account				Witness account		
		a	*b*	*?*		*a*	*b*	*?*		*a*	*b*	*?*
Guilty	prior	Likelihoods			Σ	Joint probabilities			Σ	Posterior probabilities		
A	0.30	0.80	0.10	0.10	1	0.24	0.03	0.03	0.30	0.71	0.12	0.08
B	0.10	0.10	0.50	0.40	1	0.01	0.05	0.04	0.10	0.03	0.19	0.10
?	0.60	0.15	0.30	0.55	1	0.09	0.18	0.33	0.60	0.26	0.69	0.82
					Σ	0.34	0.26	0.40	1	1	1	1

There are a number of practical applications for Bayes' theorem (French and Smith 1997). A few of them are listed in Table 7-3.

Table 7-3: Applications for Bayes' theorem

Problem	Situation	Source of information
Bring a new product to the market?	Demand	Test market
Medical diagnose	Type of desease	Examining the patient
Drill for oil?	Amount of oil	Seismic test
Cause of the plane-crash?	Pilot error, malfunction, sabotage	Investigating the wreck

The application of Bayes' theorem often leads to surprising and counter-intuitive results. Take the following example from Anderson (1989, pp. 274 f.). You go for a cancer check-up. In 5% of the cases, the test shows false results (this figure is the same for false positives and false negatives). You are told that your test is positive. Most people would now be worried sick and believe, with 95% certainty, that they have cancer. Their error is not to have taken the prior probability into consideration. The value of 95% refers to the likelihood $p(\text{pos. test} \mid \text{cancer})$. The relevant probability for you, however, is the posterior probability $p(\text{cancer} \mid \text{pos. test})$. As you know by now, when calculating a posterior probability according to Bayes' formula, you must also include the prior probability. Let's suppose yours is a rare type of cancer which only occurs in one of 10,000 people of your age. If there is no other reason for you to suppose anything but a 0.0001 prior probability, your posterior probability is

$$\frac{0.0001 \cdot 0.95}{0.0001 \cdot 0.95 + 0.9999 \cdot 0.05} = 0.0019,$$

that means less than 1:500. Particularly through its ability to correct false intuitive probabilities, Bayes' theorem becomes an invaluable prescriptive tool.

7.5 Biases in the generation of subjective probabilities

7.5.1 Introduction

Judgments concerning states of nature and thus the process of generating subjective probabilities for these states or events are based both on available data and on inferences. A number of possible biases that could influence the probability judgments can be identified and classified roughly into three categories:

1. Incomplete or inappropriate data base,
2. Incorrect processing of probabilities,
3. Insufficient critique of own judgment.

In this chapter we will attempt to give a brief outline of the most important problems. A large part of the relevant literature can be found in Gilovich et al. (2002). A very good overview is provided by Dawes (1988), Bazerman and Moore (2008) as well as Jungermann et al. (1998). For a quite entertaining discussion about (seemingly) quaint stories from the world of (im-)probabilities, see Dubben and Beck-Bornholdt (2006).

7.5.2 Incomplete or inappropriate data base

Availability

Unlike a computer, a human decision maker does not have full and constant access to all stored information. Some things "cross one's mind" faster than others do, they are more readily available. Tversky and Kahneman (1973) described *availability* as a heuristic that is used when estimating frequencies or probabilities. The easier one can find examples of a specific event, the higher the frequency or probability we attribute to this event. Availability is helpful for such estimates since events that are likely are usually more available than unlikely ones. But, on the other hand, availability is influenced by more than just frequency of occurrence (Tversky and Kahneman 1973; also see Fischhoff et al. 1978). What other factors influence availability? Especially striking events are remembered more easily than those that do not attract attention. They are more readily available, which leads to an overestimation of their frequency of occurrence. According to Dawes, this can explain the common belief that one usually ends up in the slowest queue in the supermarket or that it won't rain when you have an umbrella with you. In general, more recent events are also more available than those that occurred a longer time ago. For example, news about five train accidents within the last two weeks can influence our perception of the riskiness of travelling by train.

Hindsight

If a rather unlikely event has occurred against the odds, we tend to believe that we knew from the beginning this would happen. When asking people ex post about the subjective probability they would have attributed to that event, their answers are higher than the probabilities they attributed to the event before it occurred (Fischhoff 1975). This bias could be motivated by the wish to appear as a competent judge. However, the recently gained knowledge could also have been merged with previous knowledge in a way that makes it impossible to reconstruct one's original level of knowledge. A consequence of such biased memories can also be the systematic overestimation of one's own forecasting ability (see the discussion on overconfidence in chapter 7.5.4).

Insufficient sample size

Humans tend to base their judgments on excessively small samples, often even just on individual cases. "It was a very wise decision to go into the industry instead of staying at the university" is a statement that can neither be proven nor disproven. Wherever uncertainty is involved, the successful outcome of a decision does not mean that the decision was wise as well. Everyone knows at least one smoker aged 90 or above, but this should not lead you to draw conclusions about the risks of smoking. Using an example, Bazerman and Moore (2008) show how the advertising industry can benefit from insensitivity towards sample sizes: "Four out of five dentists suggest ..." does not indicate how many dentists have actually been surveyed.

7.5.3 Inappropriate processing of probabilities

Representativeness

Assume that a shop owner wants to hire a temporary cashier. One of the applicants is unshaven and wears jeans with holes in them. These stimuli lead the shop owner to classify him as "lazy student". With this classification comes additional (supposed) information: such people live at the expense of others, are passive, work-shy, unreliable and untidy.

Being unshaven and wearing jeans with holes in them may be representative of lazy students, but not *only* of them. Other – hard-working and ambitious – students can also have these attributes. Because of the similarity that the applicant has to "typical" members of the category "lazy student", the probability of his actually belonging to this group is overestimated (Tversky and Kahneman 1974).

Insensitivity to base rates

This is a consequence of the representativeness heuristic. Assume that 10% of all students belong to the group of "lazy students". If the shop owner had not seen the applicant, he would have had to assume a prior probability of 10% for the event "lazy student". Having seen the applicant, he neglects this *base rate* completely. For his evaluation, it is no longer relevant whether the share of "lazy students" had been 5% or 50% of the whole student population.

Neglecting prior probabilities can have disastrous consequences, for example, if entrepreneurs neglect the base rate of failure when evaluating the likelihood of success of their startup, or if newlyweds believe that they can do without a marriage contract as they think the average rate of divorces is irrelevant (Bazerman and Moore 2008).

Reversing conditional probabilities

The shop owner from our example unknowingly confuses two conditional probabilities: the probability that a "lazy student" appears down-and-out with the probability that a student appearing down-and-out is in fact lazy.

The key to avoiding this mistake is Bayes' Theorem. The posterior probability of being confronted with a lazy student is

$$p(G \mid y) = \frac{p(G) \cdot p(y \mid G)}{p(G) \cdot p(y \mid G) + p(\text{not } G) \cdot p(y \mid \text{not } G)},$$

where y denotes "down-and-out appearance" and L denotes "lazy student". If we assume the prior probability for "lazy student" to be $p(L) = 0.10$ and the likelihoods $p(y|L) = 0.25$ and $p(y|\text{not } L) = 0.05$, we can calculate a posterior probability of approximately 0.36 for "lazy student".

This "reversion of conditional probabilities" that you have already seen in section 7.4 is very common and sometimes has serious consequences. The following example is taken form Dawes (1988). According to a newspaper report, a US doctor performed pre-emptive breast amputations on 90 women belonging to a "high-risk group". The risk was determined from mammographic patterns. The reasoning behind the amputations was that 93% of all breast cancer cases belonged to the high-risk group. Thus, it holds that $p(hr|C) = 0.93$, where C stands for cancer and hr for high-risk. Considering the frequency of breast cancer for the basic population, which the doctor considered to be 0.075, as well as the fraction belonging to the high-risk group (57%), we can write Bayes' theorem in the following form

$$p(C \mid hr) = \frac{p(C) \cdot p(hr \mid C)}{p(hr)} = \frac{0.075 \cdot 0.93}{0.57},$$

The probability that a woman belonging to the high-risk group indeed gets breast cancer is hence only 12.2%.

A common argument from the discussion about drug prohibition laws is that "soft" drugs like hashish or marihuana are starter drugs and lead to further addiction to drugs such as heroin or similarly dangerous drugs. This is sometimes supported by statistics claiming that a high percentage of heroin addicts used to take soft drugs. Starting with the notion that prior hashish consumption is common for a later heroin addiction, this leads to the wrong conclusion that a subsequent heroin addiction is common for people that take hashish. (A second reasoning error in this context is related to people confusing statistical and causal correlations – somebody ending up as a heroin addict does not necessarily do that *because* he used to do hashish.)

Thinking in scenarios and the multiplication rule

Psychologists observe an intuitive violation of the multiplication rule in people having to evaluate probabilities of conjoint events. Assume for simplicity that the two sets of events, $E_1 = \{A, \text{not } A\}$ and $E_2 = \{B, \text{not } B\}$, are independent of each other. There are four possible combinations ("scenarios"): $\{A, B\}$, $\{A, \text{not } B\}$, $\{\text{not } A, B\}$ and $\{\text{not } A, \text{not } B\}$; they are displayed in Figure 7-11.

The joint probability of A and B has to be calculated using the multiplication rule: $p(A, B) = p(A) \cdot p(B)$. Based on intuition, this probability will be systematically overestimated. Oftentimes, people even estimate the conjoint probability to be greater than one of the two single probabilities, which is logically impossible. Remember the "Linda example" from Chapter 1? Here is another one: The case of a 55-year old woman who suffered a pulmonary embolism was described to internists. The doctors were asked to state the probabilities that the woman suffered from

- Shortness of breath and partial palsy,
- Pain from pleurisy,
- Unconsciousness and quickened pulse,

- Partial palsy,
- Hemoptysis.

91% of the 32 internists believed that the combination of a likely symptom – shortness of breath – with an unlikely one – partial palsy – would have a higher probability than the unlikely symptom by itself.

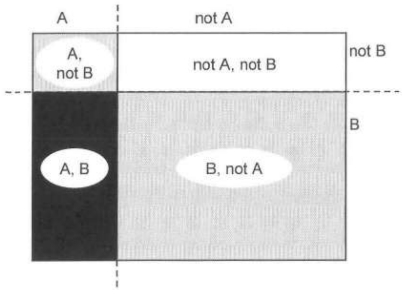


Figure 7-11: Four possible scenarios

Experiments with one fairly unlikely (A) and one fairly likely event (B) have demonstrated that the joint probability is estimated to be higher than that of the unlikely event, but lower than that of the likely event (Dawes 1988, p. 132). Thus, it is intuitively assumed that $p(A) < p(A,B) \leq p(B)$. The correct answer would obviously be $p(A,B) < p(A) < p(B)$.

An explanation of this common effect could be that humans always consider explicit scenarios that are composed of individual events and that these scenarios seem more credible than the abstract, individual events. The more details are included in the scenario, the more credible the scenario seems, even though the probability in fact decreases more and more. This mistake not only affects future events but also impacts the assessment of uncertain historical events or criminal cases (Tversky and Kahneman 1982, p. 98).

7.5.4 Insufficient critique of one's own judgment

Anchoring and adjustment

When estimating or determining a figure without a clear notion of its true level, one usually starts with a more or less random value. This starting point is then modified by further thinking, information or discussion. The starting value serves as an anchor; deviations from it are called adjustments. The literature shows that the anchor has a disproportionately high impact on the final result, i.e. that the adjustments are often insufficient. He who chooses the anchor basically determines the result. This is for instance exploited by personnel managers who ask applicants for their previous salary before making an offer.

This bias is not restricted to the task of probability estimation but is very relevant in this field. When estimating probability distributions of a numerical variable, the "average" or current value usually serves as an anchor. If, for example, you are required to give a probability distribution of the \$-€ exchange rate at the end of this year, you will most likely start with today's rate - if you happen to know it. If you do not know it, you will scan your memory for a plausible starting point. Based on this anchor, you will then try to determine the highest possible deviations (both upwards and downwards) from that rate, based on further assumptions. Laboratory experiments have produced overwhelming evidence supporting the assumption that the probability distributions of numerical variables tend to be too close to the anchor value (Lichtenstein et al. 1982). The probabilities of both extremely high and extremely low values are thus underestimated.

One way of determining whether this bias is present is to check whether the probability distribution is congruent with the rest of your beliefs. If you believe that the US-Dollar will under no circumstances fall below €0.5 or rise to more than €2, you should be ready to bet all of your assets on the event that these boundaries will not be broken either way. If you are not willing to do this, your probability distribution for the exchange rate would not be correct.

A second approach that is only feasible in very specific cases is calibration, i.e. comparing your statements and estimates to current events. A single probability statement by itself cannot be tested empirically. Assume, however, you are forecasting stock market movements of the following day for a newspaper. This forecast would concern the median as well as the 25%- and 75%-quantiles of the German stock index (DAX). After a couple of years of forecasting you would have several hundreds of your probability statement and could, for example, check how often the actual movement was above or below your forecasted median. Ideally, the actual DAX-value should have been below your 25%-quantile statement in 25% of cases, below your stated median in 50% and below your 75%-quantile statement in 75% of the cases; in that case, you would be perfectly calibrated. But if, for example, you found that the actual index was below your stated median in 80% of the cases or above your 75%-quantile statement in 40% of the cases, your calibration would be much worse and the newspaper should start looking for new staff to do the forecasts.

To prevent the tendency of giving too narrow a distribution, it is advisable not to start with the median when trying to determine a probability distribution so as to avoid an early anchoring. Neither should you attempt to determine the upper or lower boundary right at the beginning; starting with the 5%- and the 95%-fractile seems to be a more favorable approach.

Another systematic bias that can be caused by the anchoring and adjustment heuristic and which leads to problems when eliciting probability distributions is called partition dependence (Fox and Clemen 2005). If a probability distribution for a continuous variable (e.g. value of the DAX at the year-end) is needed, a common way is to ask experts to allocate subjective probabilities to different intervals (e.g. [below 4000 pts], [4000 pts, 5000 pts], [5000 pts, 6000 pts], [above 6000 pts] and subsequently derive the distribution you are looking for. The experts would typically use a close-at-hand anchor and distribute the probability mass equally over all given intervals. In our example, this would lead to an anchor of 25% for each of the four DAX-intervals. Insufficient adjustment in the next step would result in systematic biases (overestimations of rather unlikely intervals). This example once again demonstrates the importance of an adequate questioning process and of careful interpretation of the results in decision analysis. A seemingly unimportant and often arbitrarily chosen interval (or number of intervals) can have a considerable impact on the results. A prudent decision analyst should be aware of these phenomena when questioning and interpreting results (e.g. Clemen and Ulu 2008).

Overconfidence
People tend to overestimate the accuracy of their own estimates. How many states are there in Africa? Provide an interval that includes the correct number of states with a 90% probability, solve 100 similar tasks and then compare your statements with the correct values and count the number of times the correct answer was included in your interval. If about 90 of your answers are correct, you are well calibrated. Many experiments show that by far more than 10% of correct values lie outside the stated intervals. In a study by Glaser et al. (2010), financial experts had to provide detailed intervals concerning questions of general knowledge, but also specific questions about the finance sector. The error rate that – as we have just explained – should be at about 10% was in fact somewhere around 80%. The financial experts overestimated their own accuracy substantially and - interestingly - were even more overconfident and worse calibrated than a control group of students that had to provide estimates for the same questions.

This bias might be a consequence of anchoring and (insufficient) adjustment. It can have fatal or disastrous consequences, e.g. when overestimating the probability of a correct verdict in a civil process or when deciding about the launch of a new product. In 1921, US Secretary of War Baker considered the idea of a battleship being sunk by aircraft to be "insane and impossible"; he would even have agreed to stand on the bridge of a battleship being attacked by aircraft. This opinion was not founded on any type of rational information but was nevertheless believed in so strongly that no relevant information was obtained. His successor finally performed an experiment during which aircraft bombs managed to sink a battleship (Russo and Schoemaker 1989, pp. 67 f.).

We have stressed the point that probabilities should always be based on rational reasons. As the literature has shown, however, the dangers of people being overconfident with regard to their own abilities and judgments, neglecting the necessity of additional information and being guided by wishful thinking is always present.

Questions and exercises

7.1
Which interpretations of probability are these statements based upon?

(a) The probability of 8 or more pips being shown when you throw two dice is 5/12.
(b) Politician: The chances of a victory for Labor at the next election are 55%.
(c) Soccer player: Next season, my team will be promoted to the Bundesliga for sure.
(d) The probability of a child being a boy is 100/206.
(e) The probability of an accident is higher on regular highways than it is on the autobahn.

7.2
You hear the following statements of a friend and want to find out what subjective probabilities he is trying to express. Estimate the minimum and the maximum probability for each statement.

(a) I'm likely to vote for the Liberals next time.
(b) It is rather unlikely that the Millers will get back together again.
(c) I am confident that I could do my boss's job as well as he does.
(d) I feel that I'll get an A for this exam.
(e) It is impossible that I left my umbrella on the tram.

Let other people estimate the probabilities and calculate the spread between the lowest and the highest probability for each statement.

7.3
The following distribution for the anticipated costs of a certain development is given: The probability of the costs being between €200,000 and €250,000 is equal to 25%; for €250,000 - €300,000 it is 60%; and for €300,000 – €400,000 it is 15%. Within each interval, a uniform distribution is assumed.

(a) Draw the density function and the distribution function.
(b) The continuous variable "development costs" is to be discretized so that only three parameter values have to be considered. Which values would you choose and what probabilities would you assign to them? Draw the probability function and the distribution function for this case.

7.4
An estimate is needed for the number of jobs that are created in a region through the construction of a technology park. The project team in charge works out the following parameters of the probability distribution:

- at least 100 jobs are created.
- the 20% point is at 130, the 40% point is at 150, the 60% point is at 165, the 80% point is at 220 jobs.
- the maximum that seems possible is 400 jobs.

(a) Draw the distribution function by combining the mentioned points linearly.
(b) What is the probability of creating more than 200 new jobs?
(c) What is the probability of creating between 150 and 180 jobs?

7.5
Before a supplier grants a commercial loan to Loose Ltd., he tries to get an idea of the company's creditworthiness. Based on the information he already has, he assumes a 90% probability that he will not incur a loan loss.

To be certain, he asks his principle bank. Based on experience, the bank classifies 95% of the good customers as "creditworthy" and mistakenly classifies 5% of them as "not creditworthy". For customers who turn out later not to be creditworthy, the bank had given a positive statement in half of the cases and a negative statement in the other half.

The bank warns about extending credit to Loose Ltd. Based on this information, what is the probability that the supplier will later have to write off the loan?

7.6
The computer network of the WORM AG is suspected to be infected by a new and extremely dangerous virus. The corporation's IT expert judges the prior probability of an infection to be just 2%—nevertheless, he decides to purchase a newly developed scanning tool that further explores the problem. Because of the extremely tricky structure of the virus, the scanning tool cannot make definite statements about the infection. The error rate is 10%, i.e. 10% of the infected networks would be declared "clean", and 10% of the clean networks would be declared infected.

(a) Use the notation I (infected) and ¬I (not infected) for the random variable "infection", and D (virus detection) and ¬D (no virus detection) for the random variable "scanning tool detects virus". Formalize the three probability statements given in the text using this notation.
(b) What is the probability that the scanning tool will claim to have found the virus (assuming that the probability judgment of the IT-expert concerning the prior probability for an infection is correct)?
(c) When the scanning tool is used to check the network, it claims to have found the virus. Given this signal, what is the probability that the network is indeed infected?
(d) When the IT-expert checks the scanning tool manual, he reads the following phrase: "… the error rate of 10 % just refers to undetected infected systems. The error rate concerning false alerts for clean systems is significantly greater…". What impact on the a posteriori probability of a virus infection does this new information have? Please argue in general terms and do not just work through a further numerical example.

7.7
In autumn 1990, Marilyn vos Savant published the following brainteaser in her column "Ask Marilyn" in the American magazine *Parade*. A candidate in a game show is standing in front of three closed doors. He knows that behind one of the doors is the big prize, and behind each of the two other doors is a car and a goat, respectively. The candidate wins the car if he opens the right door. Assume he points at door number one. The quizmaster, however, then opens door number three and a goat is seen behind the door. Then, the quizmaster asks the candidate if

he wants to stick with his choice—door number one— or if he wishes to pick door number two.

Marilyn claimed that the candidate should change his choice as the probability of the car being behind door number two was twice as big as the probability of it being behind door one. This solution caused a flood of letters to the editor over the following months and also affected other magazines that had published the same problem. Most of the writers of these letters, among whom were a lot of mathematicians, argued that Marilyn's solution was wrong. What do you think? (The problem caused the publication of an amusing popular science book about thinking in probabilities: von Randow 1992.)

7.8
The triplets, Agatha, Beatrix and Caroline, found out that one of them was to receive a car for her birthday and the other two were each to get a receive. Agatha urges her father to tell her who will receive the car but her father refuses. Then Agatha says: "If you don't want to tell me what I will get, can you at least tell me the name of *one* of us who will get a goat?"

The father accepts this deal and replies: "One goat is intended for Caroline." Agatha is pleased because her chances to receive the car are now 50-50. Or are they?

7.9
Twenty people are at a party. Estimate intuitively the probability of two or more people's birthday being on the same day.

Then, calculate the correct probability, assuming that all days are equiprobable. (Note: The sought-for probability is 1-p(no common birthday) and it holds that p (no common birthday) is equal to $364/365 \cdot 363/365 \cdot 362/365 \cdot \ldots \cdot 346/365$.)

7.10
Dawes (1988, p. 75) cites the following extract from an article published in 1984 in the magazine *Management Focus*:

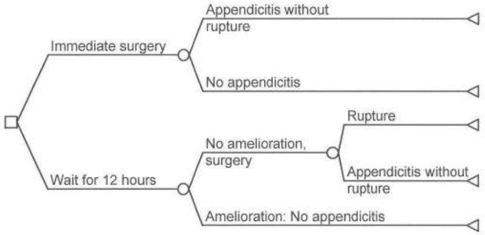
> Results of a recent survey of 74 chief executive officers indicate that there may be a link between childhood pet ownership and future career success […] Fully 94% of the CEOs, all of them employed within Fortune 500 companies, had possessed a dog, a cat, or both, as youngsters... […]The respondents asserted that pet ownership had helped them to develop many of the positive character traits that make them good managers today, including responsibility, empathy, respect for other living beings, generosity, and good communication skills.

(a) Can the stated thesis be justified by the survey results?
(b) Assume that 70% of all managers had a pet when they were a child and that 10% of all managers achieve a position as high as the managers in the survey. Can the thesis that pets form one's character be confirmed?

Case Study: Immediate appendix surgery?

Source: Clarke (1987), pp. 27-34.

The diagnosis of appendicitis is never sure. Hence, whenever signs of an inflammation appear, the problem of either operating on the patient immediately (aware of the fact that surgery might be unnecessary) or waiting arises. In the latter case, the appendix might rupture, which would lead to a much more severe condition than a simple appendicitis. The problem can be sketched thusly in a decision tree as a first approximation:



Fifty competent surgeons were confronted with the following description:

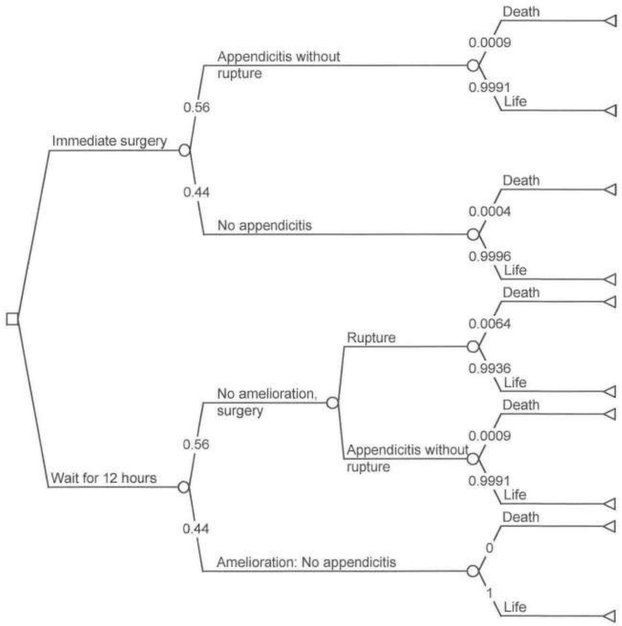
An 18-year old male with abdominal pain is taken to hospital. The results of a medical examination are somewhat in line with appendicitis, but not particularly typical. The results of laboratory test and radiograms are unclear. After reading this description, each surgeon had to decide if it was advisable to operate on the patient immediately or to wait for another 12 hours and perform surgery only if the patient's condition had not improved.

Apart from this "intuitive" decision, each surgeon was asked to assign subjective probabilities and utilities to the decision tree above.

The prior probability of appendicitis is known; also, the author could draw on the likelihoods for the results of different examinations. By application of Bayes theorem, it was hence possible to calculate the posterior probability of appendicitis. In the given case, the author concluded a probability of 0.56 for appendicitis. (The estimates made by the surgeons varied between 0.07 and 0.98!) The probability that an inflamed appendix would rupture was assumed to be 0.06.

Minimizing the probability of death served as the objective. This probability is 0.0009 for the removal of an inflamed appendix and 0.0004 for the removal of a healthy appendix. It rises to 0.0064 for the surgery of a ruptured appendix.

The decision tree was extended as follows:



Assuming these probabilities based on statistical data, immediate surgery is the (slightly) superior alternative because it yields a probability of death of 0.00068 instead of the probability of 0.00069 associated with waiting. 59% of the surgeons questioned intuitively advocated immediate surgery. Of the remaining 41%, another 90% would have come up with the "correct" result had they had worked out the alternative with the highest utility given their subjective assumptions concerning probability and utility.

Apart from the appendix-problem, the doctors were confronted with five other hypothetical problems. As a result of the survey, the author states that methods of decision analysis can significantly improve medical decision making.

Chapter 8: Simulation of an objective variable's probability distribution

8.0 Summary

1. Knowledge of the corresponding probability distributions is necessary for rational decisions between alternatives with uncertain outcomes..
2. An objective variable's distribution can be specified directly only in simple cases. If there is more than one uncertain factor influencing the objective variable, it is sensible to estimate the distributions of the influencing factors first and then derive the distribution of the objective variable afterwards.
3. It is often not possible to determine the resulting objective variable's distribution analytically. Using simulation as a tool, however, one can approximate this distribution with sufficient accuracy.
4. Today, such simulations can be executed easily with common spreadsheet programs.
5. Simulations become more complex if dependencies exist between the influencing factors that need to be considered in the simulation model.

8.1 Basic principles of simulation

In order to make an optimal decision in the presence of uncertain expectations, it is generally necessary to know the probability distributions of the outcomes resulting from each possible alternative; in simple cases, the decision maker can himself specify these distributions. Remember the publisher's example in Section 2.7.2 the publisher estimated the probabilities for different levels of demand and hence directly obtained the profit distribution for each alternative.

Usually, the circumstances are more complex. The probabilities of the outcomes are not only influenced by one but instead by several factors. As we have pointed out earlier, it is advisable in most cases to determine the probabilities for such event combinations by decomposition. The decision trees and cause trees introduced in Chapter 2 were based on this idea. For instance, the probability of radioactive leakage as a result of a reactor accident is determined by calculating the probabilities for every possible cause of interference and combining them by the rules of probability calculus.

Let us consider another example: Imagine that the unknown distribution function of the overall sales volume is required. The overall sales volume results from sales generated in many regions and within different industries. For each of these sub-segments, different experts exist who are able to evaluate the associated sales development. It hence seems reasonable to determine the sales distributions within the sub-segments and obtain the distribution of the overall sales volume by aggregation.

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_8, © Springer-Verlag Berlin Heidelberg 2010

Why do we want to know the distribution of the objective variable? The distribution contains all the available information regarding the possible outcomes of the objective variable and the corresponding probabilities. For instance, it yields the lower and upper boundary for the objective variable; we can also see which value is exceeded with a probability of, for example, 95%. Furthermore, we can calculate the average outcome (expected value), the variance or other moments of the distribution. The graphical or tabular illustration of the distribution function allows for improved insights into the risks and chances of different developments. Knowing the distribution function is a prerequisite for rational decision-making under uncertainty. In Chapter 9, we will show how the optimal alternative is to be determined in cases in which this function is known. As will become clear in Chapter 10, it is even possible to identify the optimal solution under certain circumstances even when the information regarding the distribution function is incomplete.

It is possible to *analytically* determine the distribution of an objective variable only in simple cases, and in those cases it is determined as a result of the distributions of its influencing factors. The probability distribution of the sum of two fair dice can be considered as such a simple case. The distribution is as follows

Sum of two dice	2	3	4	5	6	7	8	9	10	11	12
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

and can be determined by examining all 36 possible dice combinations occurring with equal probability and then calculating the resulting sums. Let us assume that you do not know how to determine this distribution. In this case, you could determine it empirically, e.g. by means of a repeated random experiment. This would require you to roll two dice a large number of times (say 500 times) and count the relative frequencies of the outcomes (in this case, the sum of both dice). The distribution obtained from this experiment most likely would not precisely match the theoretical one but would probably be sufficient for practical purpose.

As you can see, you can not only determine the distribution of the sum of two dice analytically but also by means of a repeated random experiment. The problem with this strategy is that even in this simple case (with an easily constructed objective variable) it would be relatively time-consuming to obtain a good approximation of the theoretical distribution. By the law of large numbers the relative frequency of an outcome will come close to its theoretical probability, with the number of trials (here throws) being sufficiently large. Such random experiments conducted in physical reality are very time-consuming and hence costly.

Simulations are often used as a tool to circumvent this time and cost problem. A simulation can be defined as a model-based experiment in an artificial environment. By analyzing the model in the artificial environment, the aim of a simulation is to generate insights which can be transferred to the real-world system. A simulation requires three main steps: formulating a suitable simulation model; mapping the model in the artificial environment; and conducting a sufficiently large number of simulation runs. The latter is required to adequately approximate the distribution of the objective variable (see above).

The starting point of any simulation is hence the formulation of a suitable model. A *simulation model* can be regarded as suitable if it is realistic enough to allow for transfer of the model-based results to the real-world system. As part of the model formulation, the distributions of the causal variables (representing the influencing factors) as well as their impact on the objective variable need to be identified. Furthermore, operational and time dependencies between the causal variables might be taken into account and adequately incorporated in the model.[1] Simulation models can be deterministic or stochastic. A model is called deterministic if all influencing factors can be assigned a certain value. However, we are mostly interested in stochastic models where the realization of at least one influencing factor is random. Furthermore, simulation models can be classified as static or dynamic. A static model maps the real-world system at only one point in time; in contrast, dynamic models depict the real-world system over time.

In the case of our dice example, a rather simple simulation model is sufficient. The two fair dice are the sole influencing factors that need to be modeled. Since the outcome of throwing a fair dice is random, the simulation model is accordingly classified as stochastic. In addition, it is a static model since the only point in time that needs to be considered in the model is the point in time immediately after the two dice are thrown; time dependencies therefore do not exist. Furthermore, the distributions of the influencing factors are known: each number is equally likely and occurs with a probability of 1/6. Since the outcomes of rolling two separate dice are stochastically independent, no dependencies between the influencing factors need to be modeled. Moreover, the impact of each influencing factor on the objective variable (sum of two fair dice) is clearly defined: the score of each rolled die serves as one summand. Consequently, the objective variable within the simulation model can be calculated by summing up the numbers of each die after they are rolled.

As the dice example shows, formulating a simulation model does not only require the specification of the relevant characteristics of the influencing factors and their impact on the objective variable but also a decision on the time sequence of events within the simulation. In the simple dice example, it is irrelevant whether you roll the two dice together or one after another. It would even be possible to incorporate only one die into the model, rolling it twice in the simulation. However, these different model formulations all have in common the fact that calculating a realization of the objective variable requires determining two die numbers beforehand. In this example, the formulation of the simulation model obviously contains some degree of freedom which can be easily understood and does not require extensive documentation. Given more complex circumstances, this is mostly no longer the case. If there are many influencing factors with various dependencies, it

[1] Given more complex circumstances, the distributions of the influencing factors as well as the operational and timely dependencies are often unknown. If in this case they cannot be derived analytically, the relevant model parameters need to be estimated.

is necessary to generate a much more detailed and well-documented formulation of the simulation model.

After formulation, the simulation model needs to be *mapped into an artificial environment* to facilitate the target-oriented experiment. Making use of computers is often the recommended strategy for doing so. In this case, the simulation model needs to be mapped into suitable software. In complex situations, tailored simulation software might be necessary; in many simpler cases, a common spreadsheet program might already be sufficient.

After mapping the simulation model in suitable software and thereby allowing for computer-aided experimenting, the model can be analyzed. Remember that the target of the simulation, and thus of this analysis, is to determine the probability distribution of the objective variable. For this purpose, it is necessary to conduct a sufficient number of *simulation runs*. As a result of each simulation run, one outcome of the objective variable is generated. After a sufficiently large number of runs and generated outcomes, the complete distribution of the objective variable becomes discernable. In addition, further increasing the number of simulation runs while changing certain model parameters (such as assumed dependencies between influencing factors or their impact on the objective variable), allows robustness tests regarding the obtained distribution. Thus, relating back to the initial dice example, rolling the two fair dice once and capturing the sum of both dice could be interpreted as one simulation run. The computer hence determines numbers for both dice randomly, calculates their sum and automatically saves the score. After 500 simulation runs, the relative frequencies of the sums are determined and illustrated as a distribution. With this step the simulation is finished and the distribution of the objective variable can be evaluated.

As pointed out by the dice example, the central element of a simulation based on a stochastic model is the generation of random numbers to adequately map possible realizations of random causal variables. In each simulation, many random samples are hence drawn. Since the random numbers generated by the computer are not the result of a true random mechanism but based on mathematical algorithms they are called pseudorandom numbers. They are created by pseudorandom number generators which are already implemented in most suitable computer programs. This computer-based numerical method of artificially generating random samples is often called *Monte Carlo simulation* or the Monte Carlo method.

The aim of this chapter is to demonstrate the use of simulation as a tool to determine the distribution of an objective variable given the distributions of uncertain influencing factors (see Hertz 1964, 1968; Schindel 1977; Hertz and Thomas 1983, 1984). For a deeper insight into the statistical basics of simulation techniques, take a look at common textbooks focusing on statistics (e.g. Mosler and Schmid 2009 as well as Mosler and Schmid 2008 and focusing particularly on Monte Carlo simulation e.g. Poddig et al. 2008). In summary, the sequential steps of a simulation are shown in Figure 8-1:

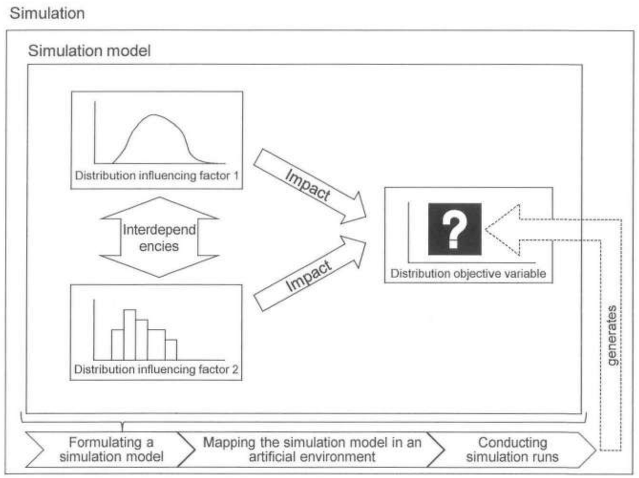


Figure 8-1: Sequential steps of a simulation

8.2 Interpretation of the simulation results

8.2.1 Economic interpretation

We will approach the interpretation of a simulation's results in two steps. We will first derive the distribution of an objective variable analytically using a simple example, allowing us to economically interpret the basic simulation approach. Second, the exact distribution calculated beforehand will then be approximated using a simulation, allowing for a methodological interpretation.

Let us begin with an example that is hardly more complicated than the one with the dice. A toy manufacturer is considering a project of producing boomerangs. Fixed costs are calculated at €4,200 per month. The sales volume is uncertain but expected to be between 1,000 and 1,500 units per month. Besides that, the profit margin is also uncertain and expected to range from €3 to €5 per boomerang. Predicting the possible scenarios, experts provide the probability distribution shown in Table 8-1.

Table 8-1: Distributions for sales volume and profit margin

Sales per month	Probability	Profit per unit	Probability
1,000	0.25	€3.00	0.30
1,200	0.65	€4.00	0.50
1,500	0.10	€5.00	0.20
	1.00		1.00

Because the number of possible profits is finite (namely nine) their respective probabilities can be calculated. Sales volume and profit margin are considered to be stochastically independent. The probability of a given profit margin, e.g. €3, thus does not depend on the realized sales volume. For each combination of sales volume and profit margin, the combined probability is simply the product of the two single probabilities. Table 8-2 presents the probabilities of all possible combinations. The cells of the matrix represent the objective variable "profit", which is given by the causal model

$$\text{profit} = \text{sales volume} \cdot \text{profit margin per unit} - \text{fixed costs}.$$

Table 8-2: Combined distribution for sales volume and profit margin with resulting monthly profits

Sales volume	Profit margin per unit					
	€3.00	(0.30)	€4.00	(0.50)	€5.00	(0.20)
1,000 (0.25)	–€1.200	(0.075)	–€200	(0.125)	€800	(0.05)
1,200 (0.65)	–€600	(0.195)	€600	(0.325)	€1.800	(0.13)
1,500 (0.10)	€300	(0.030)	€1.800	(0.050)	€3.300	(0.02)

There are eight possible outcomes for the profit (not nine, as the value 1,800 happens to appear twice by chance). By ordering the profits g_i in ascending order, each with their respective probability p_i, we obtain the profit's probability distribution in columns 1 and 2 of Table 8-3.

By consecutively summing over the probabilities, the cumulative probability distribution, or distribution function, $P_i = P(g_i)$ can be derived. This function shows for each profit the probability of *falling short of or at the most reaching* that given profit. The complement, $1-P_i$, thus shows the probability with which that profit is *exceeded*; this function is shown in column 4 of Table 8-3.

Table 8-3: Profit distribution

(1) Profit g_i	(2) Probability p_i	(3) Cumulative probability P_i	(4) Risk profile $1-P_i$
−1,200	0.075	0.075	0.925
−600	0.195	0.270	0.730
−200	0.125	0.395	0.605
300	0.030	0.425	0.575
600	0.325	0.750	0.250
800	0.050	0.800	0.200
1,800	0.180	0.980	0.020
3,300	0.020	1.000	0.000

The function $1-P_i$, - or rather its graphical representation - is known as the *risk profile*. Figure 8-2 shows the risk profile for the profit of the project. With this example, we would like to explain the meaning of the risk profile. If you want to know the probability that a given level of the objective variable will be exceeded, go upwards from that level on the x-axis until you reach the profile. For instance, the probability of breaking even is equal to 0.605 and a profit of more than €1,500 is realized with a probability of 0.2. The discontinuities are a bit more complicated. Considering for instance the probability with which a profit of 600 is exceeded, two steps are eligible, the lower of 0.25 and the upper of 0.575. The lower one is correct: the profit of 600 is *exceeded* with a probability of 25%. (In contrast, the upper step of 57.5% gives the probability that 600 is *reached* or *exceeded*.) The dotted upright lines between adjacent steps do not have a specific meaning; they are used only for illustrative purposes.

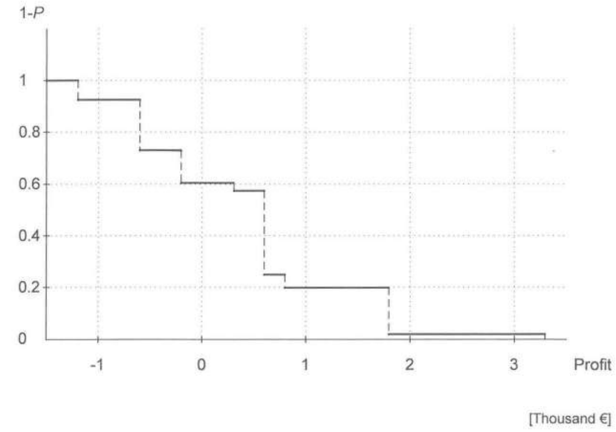


Figure 8-2: Risk profile of the profit

The risk profile characterizes the risks and the chances of the project. Additionally, the expected value (arithmetic mean) of the profit, $\Sigma p_i g_i$ can be calculated; it is equal to €402 here. One could argue that an expected value exceeding zero would lead to an approval of the project. However, this is only true under specific assumptions, which will be discussed in Chapter 9. Given that the project might yield losses, not every decision-maker would be willing to engage in it. In this chapter, we do not yet want to give a recommendation on how to decide.

8.2.2 Methodological interpretation

As shown above, the risk profile can be derived analytically - and pretty easily so - if the uncertain variable does not have many possible realizations; you do not need a simulation for this. However, we want to use a simulation in the following to show that, in this example, a simulation would just be an approximation of the analytically derivable distribution of the objective variable.

The simulation requires the following steps:

1. We randomly draw a number of the sales volume's distribution. Imagine, for example, a bin with 100 ping pong balls inside. 25 of them show the number 1,000, another 65 balls are marked with 1,200 and 10 balls show the number 1,500. The respective portions of the balls correspond exactly to the probabilities shown in Table 8.1. You now draw a ball while blindfolded and read the sales volume afterwards. Subsequently, you put the ball back into the bin. This procedure is called "drawing with replacement", en-

suring that each draw is independent of the previous one. For each draw, the number and composition of balls in the bin remains exactly the same.

2. Similarly, a number is drawn randomly from the profit margin's distribution. For this, a second bin could be used containing 10 balls, 3 of them marked with the number 3, another 5 with the number 4 and 2 with the number 5. Again, you put the ball back into the bin after each draw.
3. We have now drawn a sales volume and a profit margin. According to the causal model, we can then calculate and record the respective profit.
4. Steps 1 to 3 are repeated a total of n times, resulting in n randomly drawn profits.
5. In this example, eight different profit levels are possible. We sort them in ascending order and count how often each level is generated by the randomly drawn sales volume and profit margin. In the following, f_i denotes the relative frequency of the profit level g_i.
6. By summing over the f_i, we obtain the cumulative frequencies F_i as well as the complements $1-F_i$ which we have come to know as the risk profile.

Table 8-4 shows the relative frequencies f_i of the profits g_i as well as the corresponding cumulative functions F_i and $1-F_i$, derived with $n=25$.

Table 8-4: Frequencies of the profits for a simulation with $n=25$

Profit g_i	Frequency	Relative Frequency f_i	Cumulative relative Frequency F_i	$1-F_i$
−1,200	1	0.04	0.04	0.96
−600	5	0.20	0.24	0.76
−200	1	0.04	0.28	0.72
300	0	0.00	0.28	0.72
600	10	0.40	0.68	0.32
800	1	0.04	0.72	0.28
1,800	7	0.28	1.00	0
3,300	0	0.00	1.00	0

The interpretation of a simulation as an approximation of the true distribution can now be illustrated by use of the relative frequencies f_i as approximations of the probabilities p_i. Correspondingly, the F_i approximate the P_i. The approximation improves with the number of simulation runs n. Figure 8-3a shows the risk profile derived from the example above with $n=25$ and Figure 8-3b shows the risk profile for $n=250$. The profiles from the simulation are shaded; the exact profile from the calculation is represented by the solid line. One can see that the approximation improves with the number of simulation runs; the shaded profile better fits the solid line.

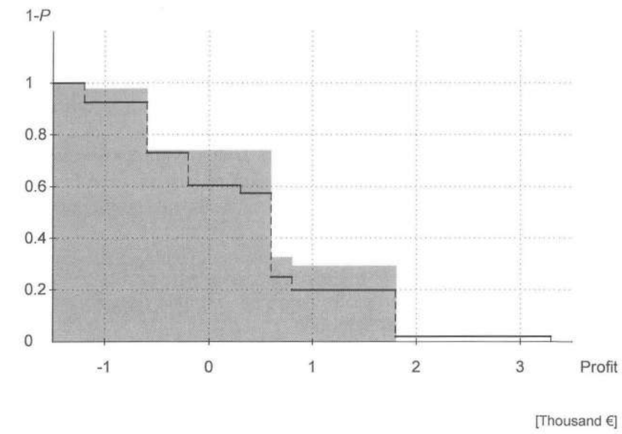


Figure 8-3a: Simulated risk profile for $n=25$ and exact risk profile

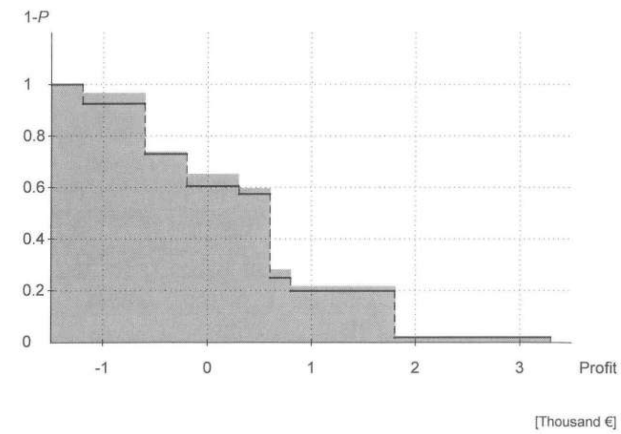


Figure 8-3b: Simulated risk profile for $n=250$ and exact risk profile

Figure 8-3a/b shows the results for a specific set of $n=25$ and $n=250$ simulation runs (and hence, randomly drawn numbers). If you were now to repeat the simulation with the same number of runs, you would most likely obtain a slightly different result due to a different set of random numbers being drawn. The variance in the result, however, decreases for an increasing number of simulation runs as the overall approximation improves. A larger number of simulation runs usually coincides with higher costs for the simulation (e.g. higher required computational power for more complex simulations). Consequently, one has to balance the desired degree of precision with the acceptable costs for conducting the simulation.

8.3 Conducting a simulation

8.3.1 Transformation of a random number into a realization of the causal variable

As we have seen, random realizations have to be drawn for the causal variables, according to their respective probability distributions, in order to simulate the distribution of the objective variable. However, using balls or slips of paper to randomly draw the numbers is certainly not efficient. In this section, we will demonstrate how drawing a realization of a uniformly distributed random variable and the subsequent transformation into a causal variable's outcome can be performed using a computer program.

Let us start with the discrete case. Consider first the variable "sales volume" from our boomerang example (see Table 8-1), which we will denote by x and which has possible outcomes of x_i = 1,000; 1,200; 1,500. To these three outcomes we have assigned respective probabilities of p_i = 0.25; 0.65; 0.10. Common spreadsheet programs usually have built-in functions for generating random numbers, which are usually distributed uniformly across the interval between zero and one. A given realization of the uniform distribution can easily be transformed into the corresponding value of the causal variable's distribution using the distribution function; this transformation will now be explained. The values P_i of the distribution function represent the probabilities of the events $x \leq x_i$. It follows that

$$P_i = p(x \leq x_i)$$
$$P_{i-1} = p(x \leq x_{i-1})$$

and hence in the discrete case

$$P_i - P_{i-1} = p(x = x_i). \tag{8.1}$$

When drawing a number z that is uniformly distributed between zero and one, the probability that the realization lies in the interval between P_{i-1} and P_i is equal to P_i-P_{i-1}, which is equivalent to $p(x=x_i)$ according to (8.1). Consequently, every time the random number lies within this interval, we can take x_i as the relevant realization of the causal variable.

Following this, we obtain x_i with its according probability. Relating back to our boomerang example, for random numbers between 0 and 0.25 we take the outcome to be equal to 1,000, for numbers between 0.25 and 0.9 the outcome is 1,200 and for numbers between 0.9 and 1 the outcome is 1,500. Using Microsoft Excel,

the corresponding function to generate uniformly distributed random numbers between zero and one is called "*rand()*".

The left panel of Figure 8-4 illustrates this; it shows a distribution function for the sales volume. Every time a random number z between 0 and P_1 is drawn, 1,000 boomerangs are sold. If a random number lies between P_1 and P_2, 1,200 units are sold and if it lies between P_2 and P_3, 1,500 units are sold. For a sufficiently large number of simulation runs, we end up with approximately the given probability distribution of x.

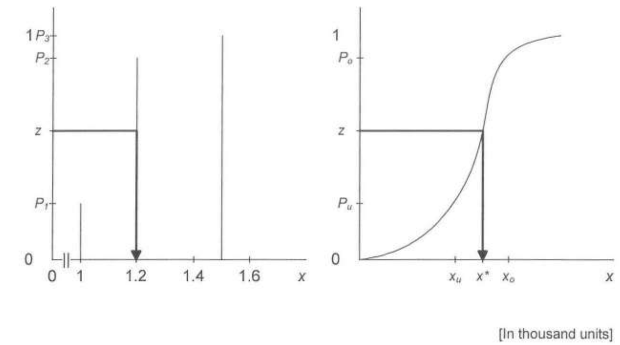


Figure 8-4: Transformation of a uniformly distributed random number between zero and one into a discrete (left panel) or continuous (right panel) causal variable with given distribution

In the continuous case, the procedure is very similar. For two outcomes x_u and x_o with $x_o \geq x_u$, it follows that

$$P_o - P_u = p(x_u < x \leq x_o). \tag{8.2}$$

The relative frequency with which the uniformly distributed random number lies in the interval $[P_u, P_o]$ can be seen as an approximation for the proability with which x lies in the interval $[x_u, x_o]$. Assume that z is drawn as the random number; we then assume the outcome x^* of the causal variable for which it holds that $P(x^*) = z$. To rephrase, we use the random number as the argument for the inverse of the distribution function:

$$x = P^{-1}(z). \tag{8.3}$$

This is illustrated in the right panel of Figure 8-4. Consider a random number z. From that number, you go horizontally to the right until the distribution function is reached. The outcome x^* for the causal variable can be obtained by vertically going down until the x-axis is reached.

Assume that the causal variable follows a standard normal distribution. The outcome could then be generated in Microsoft Excel using the functional linkage "*normsinv(rand())*". As shown above for the discrete case, the inner function generates a random number distributed uniformly between zero and one. This number is then inserted as the argument for the inverse of the distribution function. As a result, one obtains a random outcome of the standard normally distributed causal variable.

8.3.2 Flowchart

The procedure of a simulation can be illustrated by the following flowchart:

1. Conducting a single simulation run:
   - For every causal variable i ($i=1,2,...,k$), a random number z_i distributed uniformly between zero and one is generated.
   - According to the functional forms for discrete and continuous distribution functions P_i, the outcome x_i of the i-th causal variable is calculated for $i=1,2,...,k$.
   - All outcomes x_i ($i=1,2,...,k$) are inserted into the causal model and the outcome of the objective variable $y=y(x_1,x_2,...,x_k)$ is be obtained.
2. Determining the distribution of the objective variable:
   - After a given number of n simulation runs, the outcomes drawn for the objective variable are sorted in ascending order. Assume that outcome y_j appears m_j times, $1 \leq j \leq q$ as well as $y_1 < y_2 < ... < y_q$.
   - The relative frequencies $f_j = m_j/n$ and the cumulative relative frequencies $F_j = \Sigma_{l=1}^{j} f_l = p(y \leq y_j)$ are then calculated.
   - The risk profile can then be derived:

$$1-P_j = \begin{cases} 1, \text{if } y < y_1 \\ 1-F_j, \text{if } y_j \leq y < y_{j+1} \quad (\text{for } j=1,2,...,q-1) \\ 0, \text{if } y_q \leq y \end{cases} \tag{8.4}$$

8.3.3 The boomerang example in the continuous case

To illustrate the aspects discussed so far and in connection with Section 8.4, we will extend the boomerang example to the continuous case. The usage of normally distributed causal variables is very common, which is why we use such a setting in the next section to model dependencies among variables. Normally distributed variables are often used to simulate market prices (stock prices, exchange rates, oil price).

At the beginning of the boomerang example, the profit margin per unit was itself a causal variable. However, the profit margin comprises different single components which we will consider separately in the following. As single components we will take the sales price per unit, the costs for the raw material per unit (especially for plywood), the production costs per unit (especially for energy) as well as other variable costs per unit. The other variable costs are assumed to be €3 per unit. Fixed costs remain at €4,200 per month. Since the sales price can be set by

the manufacturer (and thus is not uncertain), we assume it to be €15 per unit. The objective variable is still the profit per month which follows from the causal model

$$\begin{aligned} \text{profit} = \text{sales volume} \cdot (&\text{price per unit} - \text{raw material costs per unit} \\ &- \text{production costs per unit} - \text{other variable costs per unit}) - \text{fixed costs}. \end{aligned} \tag{8.5}$$

The causal variables sales volume, costs of raw material and costs of production will be modeled as continuous variables and more precisely as normally distributed variables. This requires that for each variable, the mean and the standard deviation have to be specified. Table 8-5 shows all the necessary parameters. Figure 8-5 comprises graphical presentations of the density function and the distribution function for the three causal variables.

Table 8-5: Data for the extended boomerang example

		Mean	Standard deviation
Causal variables			
	Sales volume	1.200 units	250 units
	Costs of raw materials per unit	€5.00	€1.00
	Costs of production per unit	€2.00	€0.25
Fixed data			
	Fixed Costs	€4,200.00	
	Other variable costs per unit	€3.00	
	Price per unit	€15.00	

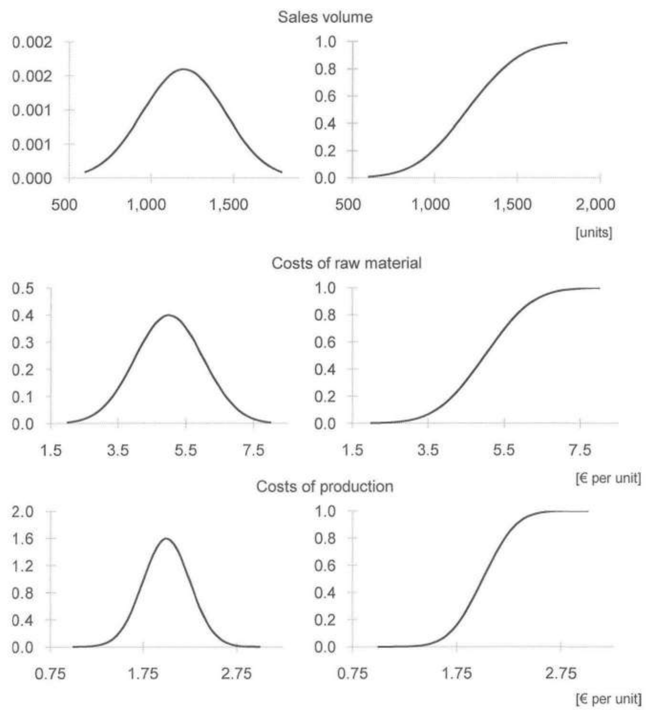


Figure 8-5: Density functions and distribution functions for the causal variables

We now have all data to start the simulation. Let us recall the first step of the flowchart which is conducted once for every simulation run. In this example, we use $n = 250$ simulation runs. For each run and each causal variable, we draw a random number. As we initially do not take any dependencies among the causal variables into account, the random numbers are drawn independently for every causal variable and are subsequently inserted into the inverse of the respective distribution function, leading to a random outcome of the variable. As shown above, one could use "*normsinv(rand())*" in Microsoft Excel for standard normally distributed variables. However, in this example, the variables are not standard normally distributed but instead normally distributed with individual mean and standard deviation. Because of this, the values of the standard normal distribution have to be adjusted. This can be done by multiplying the number from the standard normal distribution with the standard deviation (for sales volume: 250) and by subsequently adding the mean (for sales volume: 1,200). Microsoft Excel provides a way to directly combine the drawing of a random number with the transformation into a normally distributed number:

„*norminv(rand();<mean>;<standard deviation>)*"

The function call for the sales volume would thus be

„*norminv(rand();1200;250)*"

With regard to the modeling of dependencies, which will be discussed in section 8.4, one however needs the manual transformation.

For every simulation run we obtain an outcome for the sales volume, the costs of raw material and the costs of production. Inserting these numbers into the causal model in (8.5), we obtain the profit per month. Over all simulation runs we therefore get $n = 250$ simulated profits. In step 2 of the flowchart we then sort the profits in ascending order and calculate the cumulative relative frequencies. Based on these, we can then plot the risk profile as shown in Figure 8-6.

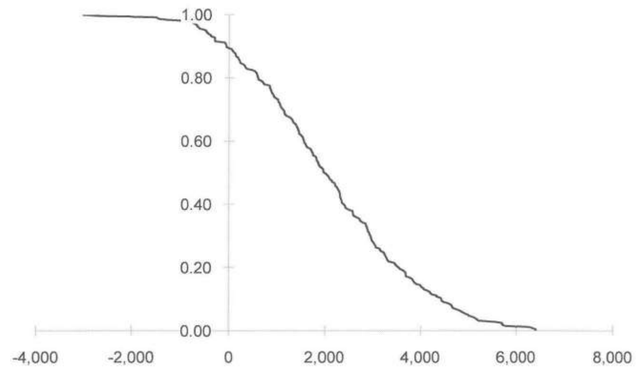


Figure 8-6: Risk profile for $n = 250$ simulation runs for the extended boomerang example

By means of the risk profile, one can see that a profit exceeding zero is achieved in 90% of the runs. The risk profile thus looks good in the first view as the loss probability amounts to only about 10%. In some cases, examining only the risk profile (or comparing different risk profiles) can already be sufficient to lead to reasonable decisions. In most cases, however, the risk profiles are not as unambiguous as required and thus do not qualify as a basis for a decision. Chapter 10 will come back to this issue.

8.4 Modeling interdependencies between causal variables

When stochastic dependencies between causal variables have to be considered, simulations may grow complex. However, interdependencies are frequent in reality. There are different ways to handle them. Drawing on the boomerang example, we will now discuss three ways of doing so.

8.4.1 Conditional probability distributions

Particularly in the case of only a few discrete values of the causal variables, it may prove practicable to work with conditional probabilities. For this purpose, the causal variables are repeatedly simulated in the same order. The probability distribution of each variable may be influenced by the value of an already simulated causal variable. In the boomerang example, you could start off by defining different values and probabilities of occurrence for the costs of raw materials. For the costs of production, there would then be multiple probability distributions, each corresponding to one parameter value of the costs of raw materials. The raw material costs and the production costs are presumably positively correlated, for example because of common exposure to inflation. The conditional probability distributions for the costs of production should therefore yield higher "expected" costs of production if costs for raw materials are already high, and vice versa. If a relatively high value of raw material costs is drawn in the course of a simulation, the costs of production will be evaluated using a probability distribution which is pushed further to the right, i.e., has a higher expected value. This way, larger costs for raw materials will tend to coincide with high costs of production.

8.4.2 Accessing causing variables

Theoretically, it would be superior to trace back the interdependencies between causal variables to one or more *causing variables*. If both raw material and production costs depend on the general level of prices then inflation should be modeled as a variable in the simulation. The variable costs could then be a function of inflation, or the influence of the inflation on both cost components could be explicitly modeled. Ultimately, all models could be traced back to a set of causing variables which could legitimately be assumed as being stochastically independent of each other.

In reality, only rarely will it be possible to break a model down to include only independent causal or rather causing variables; this would possibly require the simulation of various variables. Also, estimating the functional relationships between the causal variables and the objective variable would be very complex or not feasible at all. Therefore, it is often practicable to predict the relationships between the causal variables and to incorporate these interdependencies into the model. With normally distributed variables like in the boomerang example, this might be possible with reasonable effort. Since such methods are of high practical relevance, we will outline the necessary procedures in the following final part of this chapter without putting too much emphasis on mathematical derivations.

8.4.3 Modeling interdependencies using correlation matrices

In addition to the assumptions listed in Table 8-5, it holds that costs for raw materials and production are correlated with a correlation coefficient of $\rho = 0.4$. You might remember the concept of correlation from statistic classes: correlation measures the strength of the interrelation between two random variables. A positive correlation means that larger parameter values of one variable tend to coincide with larger values of the second random variable. Such a relationship was assumed for both types of costs for the boomerang example in Section 8.4.1. We will not elaborate further on how the correlation of variables can be forecast; time series analysis or expert estimates may serve as starting points for this purpose. Graph 8-7 shows how correlation affects the simulation of parameter values. Combinations of randomly drawn values for the costs of raw materials and production are depicted in a coordinate system. For Figure 8-7a, it is assumed that the two variables are uncorrelated, i.e. $\rho = 0$. The random draws result in a diffuse scatter-plot. However, in Figure 8-7b, a correlation of 0.4 is assumed and a (weak) trend can be observed, linking higher costs of production to higher raw material costs. For the purpose of simulations, the capability of generating such combinations of correlated random variables is important.

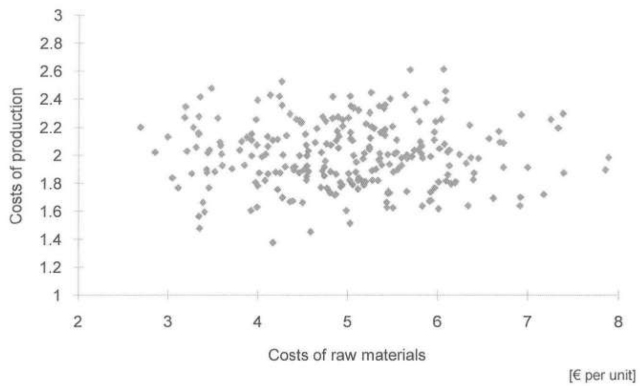


Figure 8-7a: Stochastically independent raw material and production costs with $n = 250$

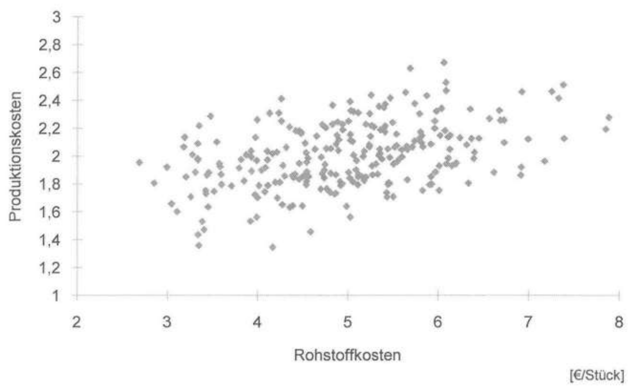


Figure 8-7b: Simulated correlated raw material and production costs with $n = 250$ and $\rho = 0{,}4$

In order to include the correlation of causal variables in the simulation model, only one additional step is required. As a prerequisite, a correlation matrix has to be prepared, containing the correlation $\rho_{a,b}$ between two arbitrary random variables a and b (one defining the column, the other defining the row of the matrix). For n random variables, we obtain an $n \times n$-matrix; with only two variables, a and b, we can set up the following simple matrix:

$$\mathbf{K} = \begin{pmatrix} \rho_{a,a} & \rho_{b,a} \\ \rho_{a,b} & \rho_{b,b} \end{pmatrix}$$

In our example, a will stand for the costs of the raw materials and b for the production costs; we hence obtain the following matrix:

$$\mathbf{K} = \begin{pmatrix} 1 & 0{,}4 \\ 0{,}4 & 1 \end{pmatrix}$$

Since each variable is perfectly correlated with itself, all diagonal elements of the correlation matrix are equal to 1. In addition, all correlation matrices have to be symmetric with respect to the principle diagonal as it holds that $\rho_{a,b} = \rho_{b,a}$.

The further procedure is explained most conveniently using matrix calculus and the respective notation.

For each correlation matrix K, there exists a triangular matrix G with the property that:

$$\mathbf{K} = \mathbf{G}\mathbf{G}^T \tag{8.6}$$

This decomposition of K is called the "Cholesky decomposition" and the matrix G is known as the "Cholesky matrix". While using matrix multiplication to prove that a given matrix G has the properties defined by (8.6) seems to be quite feasible to, it is not trivial to compute the Cholesky matrix for a given matrix K. Unfortunately, Excel does not include a standard function for calculating the Cholesky matrix - however, you will probably be able to find an application (e.g. on the Internet) that can compute the Cholesky decomposition for you.

In specialized statistical software, like the freely available *R*, obtaining the Cholesky decomposition is easily accomplished. This will require only a few commands which can be learned without having to acquire extensive knowledge of the software's other operations. In the case of the boomerang example, the Cholesky matrix can be worked out as

$$\mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0{,}4 & 0{,}9165 \end{pmatrix}$$

After the Cholesky matrix corresponding to the respective correlation matrix has been calculated, the follow-up steps are quite simple. As in Section 8.3.3, the Excel formula "*normsinv(rand())*" is used to generate a value z_i from a standard normal distribution for each random variable i. At this stage, the normally distributed random values z_i are uncorrelated. If we take each generated combination of the z_i as vectors z and multiply these vectors by the Cholesky matrix G, we obtain vectors x containing random values which reflect the desired correlation. It holds that:

$$x = \mathbf{G} \bullet z \tag{8.7}$$

After this intermediate step which imposed the desired correlation structure onto the random variables, we can continue with our procedure introduced in Section 8.3.3. The standard normally distributed values x_i have yet to be adjusted with respect to the given expected values and standard deviations. To this end, each x_i is multiplied by the standard deviation of the causal variable i and the respective expected value is added.

The whole procedure shall be exemplified by looking at one simulation run. Since the matrix G has been stated already, we can progress by generating two uniformly distributed random numbers. Assume those numbers are 0.2 for the raw material costs and 0.6 for the costs of production. Both values are converted into realizations of a standard normally distributed random variable using Excel:

$$\textit{standnorminv}(0.2) = -0.8416$$
$$\textit{standnorminv}(0.6) = 0.2533$$

As outlined above, the latter two steps may be combined into one using the Excel command "*normsinv(rand())*"; we just wanted to make the methodology as clear as possible. Subsequently, the vector of the random numbers is multiplied by the Cholesky matrix:

$$\begin{pmatrix} -0.8416 \\ 0.1045 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.4 & 0.9165 \end{pmatrix} \bullet \begin{pmatrix} -0.8416 \\ 0.2533 \end{pmatrix}$$

It follows that the simulated costs amount to:

$$\text{Raw materials: } -0.8416 \cdot €1.00 + €5.00 = €4.16,$$
$$\text{Production: } 0.1045 \cdot €0.25 + €2.00 = €2.03.$$

Since it still holds true that the sales volume is stochastically independent of the cost components, the sales can still be simulated as outlined in Section 8.3.3.

Due to the positive correlation of the costs, the risk profile for the boomerang example will change; scenarios with particularly high or low costs will become more likely. This consequence of the correlation can be seen in Figure 8-8. 250 simulation runs were conducted using the procedure described in this section. In this new analysis, the probability of a loss amounts to 18%, which is substantially more than the 10% estimated using uncorrelated random variables. How serious the consequences of ignoring correlations can be in practice was demonstrated just recently in the subprime-crisis. Banks forecasted probabilities of default without allowing for correlations and hence underestimated the probability of a scenario of simultaneously increasing interest rates, falling real estate prices and decreasing abilities of the borrowers to pay off their debt.

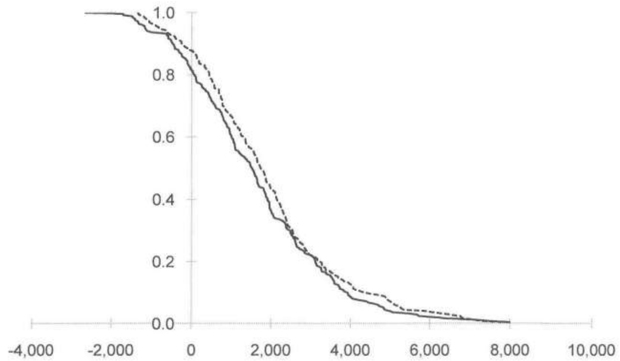


Figure 8-8: Risk profile for $n = 250$ simulation runs for the boomerang example with correlated cost components (the dotted line shows the risk profile from Figure 8-6 for reasons of comparison)

Questions and excercises

8.1

A project has the following consequences (costs in €) and probabilities. Draw the project's risk profile.

20,000	22,000	24,000	26,000
0.1	0.25	0.35	0.30

8.2

If you have a spreadsheet program with an integrated random number generator, define two uncertain independent variables which are equally distributed over the interval [10, 20] and link them multiplicatively. Do 100 draws (Hint: define an extra row for every draw!), sort the numbers in ascending order and draw the distribution function.

8.3

The director of the municipal museum, Dr Thanks, is asked by the City treasurer what number of visitors s he expects for the coming year. Dr Thanks divides the visitors into three categories: students, tourists and others.

	Lower limit	Mode	Upper limit
Students	800	900	1,200
Tourists	1,000	1,300	1,600
Others	1,500	2,000	3,000

He thinks the attendances of these three categories are stochastically independent. He approximates the distribution using a triangular distribution (table).

(a) Try to derive the distribution function of a triangular function with the parameter a (lower limit), b (mode), and c (upper limit).

(b) Dr Thanks looks up the solution to exercise (a); it is

$$P(x) = \frac{(x-a)^2}{(c-a)\,(b-a)} \quad \text{for } x \le b$$

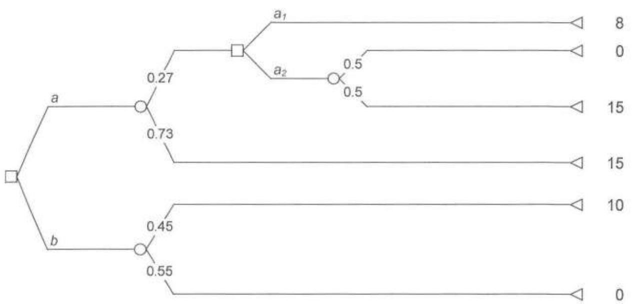
$$P(x) = 1 - \frac{(c-x)^2}{(c-a)\,(c-b)} \quad \text{for } x > b$$

Create a risk profile of the visitor numbers with the aid of these functions (and preferably a computer).

(c) The city treasurer indicates that the museum will have to be closed if the visitor numbers are below 4,500 a year. What is the probability of more than 4,500 visitors?

8.4

Three strategies are possible in the following decision tree. Calculate the risk profile for each of the three strategies.



8.5

(a) Characterize the decision situation in which a simulation seems suitable. Which preconditions have to be met?

(b) Discuss the problems of a simulation with multiple objectives.

8.6

A company's turnover is the sum of the normal distributed turnovers of the products A, B, and C. The following table depicts the parameters of each of the normal distributions. The turnovers are stochastically independent. Calculate analytically the distribution of the total turnover.

Product	Expected value	Standard deviation
A	€1 Mio.	€0.5 Mio.
B	€2 Mio.	€0.5 Mio.
C	€4 Mio.	€1.5 Mio.

8.7

Otto, owner of a fish and chips shop, is considering setting up his stall at the Send, a fair in Münster which will last for ten days. Otto will have to incur fixed costs of €2,200 (for waste disposal, electricity, etc.). Otto generates revenues by selling fried fish and as well as herringsandwiches. The realized turnover will depend strongly on weather conditions. When it is cold, people tend to buy more fried fish ; when it is warm, they tend to buy more sandwiches. When it is rainy, the overall turnover will generally be lower. Because of his fascination for statistics, Otto has gathered data for many years concerning revenues and weather dependencies. This way, he is able to make accurate estimates for his sales of fried fish and herring-sandwiches.

For the whole ten day period, he expects to sell 2,000 servings of fried fish, an amount that may vary depending on the weather. Otto assumes a normal distribution as a fairly accurate approximation with a standard deviation of 600 servings. He makes €4.50 per serving of fried fish which costs him €1.50. The herring-sandwiches cost €1 per serving for Otto who sells them to his customers for €3 per sandwich. He expects to sell 1,500 sandwiches and again assumes a normal distribution, this time with a standard deviation of 400. Since the sales volumes both depend on the amount of rain (which unfortunately is quite common in Münster), Otto assumes a correlation of 0.7 between the two sales volumes.

To Otto, it is of primary concern to know the probability that the revenues after deduction of the fixed costs will not even be sufficient to pay his employees; the personnel costs amount to €2,800. In addition, he is interested in the probability that a gross entrepreneurial profit of €2,000 (which he deems appropriate) is exceeded.

(a) Conduct a simulation to answer Otto's questions.

(b) Would it prove advantageous to Otto, if he sold the fried fish for €5? He assumes that the mean amount of sold servings would decrease by 300, with no change in the standard deviation.

(c) Otto now remembers that he usually gives away a free serving of fried fish to his friends – of which he has quite a few when working in his shop). The number of friends visiting him at his shop varies uniformly between 100 and 150 without any correlation with the overall sales volume (which does not include the free servings). How does Otto's generosity influence the objective variables mentioned above?

Case study 1: Bidding for butter

Source: Hertz and Thomas (1984).

The British supermarket chain J. Sainsbury received the opportunity to act as a bidder at the auction of a shipment. The cargo contained butter and "sweetfat", a butter-like product, in an unknown proportion. At that time, the import of butter was legally restricted. The cargo was declared as "sweetfat" but presumably included a large share of butter. How much exactly could not be estimated reliably. The customs authorities took a sample of 250 out of the total of 20,000 cardboard boxes. Each company interested in the auction was given ten of those boxes. Sainbury's sample contained eight boxes of butter and two boxes of sweetfat.

Due to considerations which shall not be discussed in greater detail Sainsbury's chief procurement officer estimated the following probability distribution.

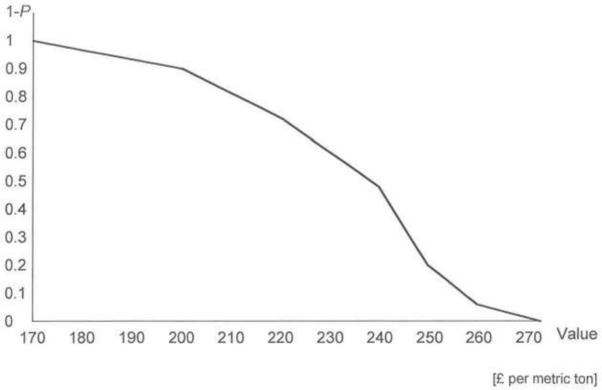
% butter	85	80	75	70	65	60	55
Probability	0.05	0.20	0.40	0.20	0.05	0.05	0.05

Both the future purchase price of butter and of sweetfat was uncertain. Again, a probability distribution was created.

Price for butter £/to	200	220	240	260	280	300
Probability	0.05	0.10	0.10	0.20	0.50	0.05
Price for sweetfat £/to	110	130	150	170	190	210
Probability	0.05	0.40	0.20	0.20	0.10	0.05

Obviously, the price for a metric ton of the cargo is $f \cdot b + (1-f) \cdot s$ with f denoting the share of butter, b denoting the price of butter and s denoting the price of sweetfat. From the given probability distributions, a risk profile of the value per metric ton can be derived. The prices for butter and sweetfat were assumed to be stochastically independent; this seems problematic. The simulation yielded an expected value per metric ton of £233.

Surely, this analysis cannot answer the question of what price Sainsbury should bid at the auction. For this, the different probabilities of bidding successfully at the auction given alternative bids have to be considered. These considerations shall not be presented in this short example. However, they lead Sainsbury to bid a price of £203 per metric ton. This turned out to be the winning bid and Sainsbury realized a profit of £40 per ton.



Case study 2: Portfolio choice

Source: *fictitious example.*

A mutual fund splits up its wealth equally into the stocks of two banks A and B. From experience, the fund management knows that the stock of bank B exhibits greater volatility than that of bank A. In fact, it is assumed that the return on both stocks over the next year will follow a normal distribution with an expected return of $\mu_A = 10\%$ for bank A and $\mu_B = 12\%$ in the case of bank B. The volatility of returns is expressed by the standard deviation of the respective normal distributions: $\sigma_A = 15\%$, $\sigma_B = 20\%$. In the past, the returns on both stocks were uncorrelated ($\rho_{A,B} = 0$), so the fund management expects this condition to hold true for the future as well. The fund's risk management is interested in the Expected Shortfall for the 5% quantile, assuming that the fund splits its assets evenly over both stocks. This Expected Shortfall is defined as the expected return given that the return is within the 5% of worst possible fund performances over the next year. However, the management does not know that the two banks are engaged in off-balance-sheet vehicles which invest in very similar, risky assets. The true correlation of stock returns therefore amounts to $\rho^*_{A,B} = 0.5$.

A simulation may be implemented to assess the future fund returns. Instead of computing the distribution of returns explicitly, returns for the stocks of bank A and bank B (X_A and X_B) may be drawn randomly from a normal distribution and then combined to yield the respective fund return. In fact, the return of the fund X_{fund} is simply given by:

$$X_{fund} = \frac{1}{2} X_A + \frac{1}{2} X_B,$$

with $X_A \sim N(0{,}1;0{,}15^2)$ and $X_B \sim N(0{,}12;0{,}18^2)$.

The simulation may be conducted using a spreadsheet program or statistical software. E.g., in Excel the function „*norminv(rand();0.1;0.15)*"can be used to obtain random returns for stock A. The formula can easily be changed to yield random returns for B by plugging in the expected value and standard deviation of B's normal distribution.

The calculation grows a little more complex if the true correlation of the two stocks is to be accounted for. Firstly, the correlation matrix K has to be formulated:

$$\mathbf{K} = \begin{pmatrix} 1 & 0{,}5 \\ 0{,}5 & 1 \end{pmatrix}$$

Secondly, a Cholesky decomposition of K is obtained as the triangular matrix G:

$$\mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0{,}5 & 0{,}866 \end{pmatrix}$$

Standard normally distributed random numbers Z_A and Z_B may then be simulated:

$$Z_A, Z_B \sim N(0;1).$$

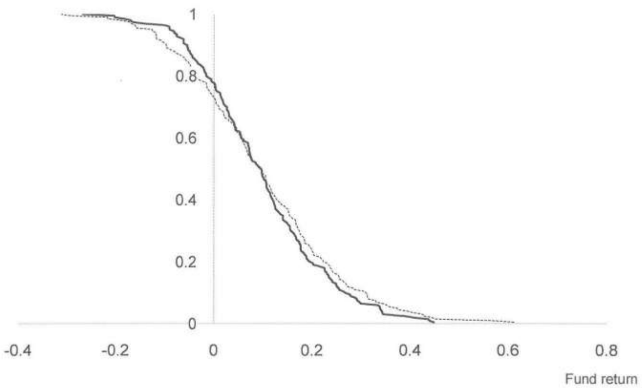
The correlated standard normally distributed random variables follow by multiplying the matrix G by the vector z containing the realizations of the random variables Z_A and Z_B:

$$Y_A = Z_A,$$
$$Y_B = 0{,}5 \cdot Z_A + 0{,}866 \cdot Z_B.$$

The resulting values are then transformed to reflect the desired normal distribution:

$$X_A = Y_A \cdot \sigma_A + \mu_A,$$
$$X_B = Y_B \cdot \sigma_B + \mu_B.$$

At this stage, the fund return can be calculated as the average of the returns of the stocks of bank A and B just as in the case of uncorrelated returns. Arranging the fund returns in ascending order, one can draw a risk profile as shown in Figure 8-10 (for $n=200$ simulation runs for the mutual fund example with correlated stock returns [dotted line] and uncorrelated returns [solid line]).



The risk profile shows that in the case of correlated returns, scenarios with very high and very low returns are more prevalent compared to the case of $\rho_{A,B}=0$. The Expected Shortfall may then be calculated as the average over the 5% of the worst realizations of X_{fund}. If the fund management knew about the true correlation, it would assume an Expected Shortfall of 23.2% according to the simulation results depicted in Figure 8-10. However, if the management relied solely on the assumptions of uncorrelated returns, it would underestimate the true risk and decide on its investment strategy based on an estimated Expected Shortfall of only 16.4%.

Chapter 9: Decisions under risk and one objective

9.0 Summary

1. For decisions under risk the theory of expected utility establishes the basis of rational action.
2. A preference regarding risky alternatives can be represented by a utility function, if the preference fulfills the axioms of "complete ordering", "continuity" and "independence".
3. There is a measure by Arrow and Pratt for a decision maker's risk attitude concerning lotteries. The measure is greater (smaller) than zero, if the utility function is concave (convex). A concave utility function is equivalent to risk aversion, a convex utility function is equivalent to risk proclivity.
4. Using the terms risk aversion or risk proclivity, it must be pointed out that aspects of risk and aspects of value are inseparably inter-connected.
5. Only in a few cases is the mean variance approach compatible with the theory of expected utility; it should hence only be applied with great caution.
6. You will learn five methods of determining a utility function: the bisection version of the variable certainty equivalent method, the quantile version of the variable certainty equivalent method, the variable probability method, the method of equal utility difference, and the trade-off method. All methods are based on the repeated evaluation of simple lotteries.
7. Consistency checks are an essential component of the methods of determining a utility function.
8. The expected utility can easily be calculated in the single-stage model; the optimal alternative is the one with the highest expected utility. In the multi-stage model, the optimal alternative is computed with the aid of the roll-back method.

9.1 Evaluation of risky alternatives

In chapters 7 and 8, several display formats for decisions under risk were elaborately presented. In particular, probability theory was introduced and it was shown how subjective probabilities can be derived from decision makers' judgments. In this chapter, we want to model a decision maker's preference in case he faces risky alternatives. We first examine the simple case that the decision problem can be represented in the form of a single-stage model, i.e. the alternatives can be defined in terms of single-stage lotteries. Multi-stage alternatives that are represented by decision trees or strategies will be discussed in more detail in Sec-

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_9, © Springer-Verlag Berlin Heidelberg 2010

tion 9.5. For the sake of simplicity, we will predominantly look at alternatives with a finite number of consequences.

If a set of alternatives consists of two risky alternatives a and b, the alternatives may be represented by the vectors $(a_1, p_1; ...; a_n, p_n)$ and $(b_1, q_1; ...; b_n, q_n)$. This notation states, that consequence a_i occurs with probability p_i. The alternatives can also be represented by a decision matrix or graphically, as in Figure 9-1.

Figure 9-1: Graphical representation of single-stage lotteries

In Chapter 5, we demonstrated how the decision maker's preference can be represented for decisions under certainty with one objective. Analogously, we will explain in this chapter how a decision maker's preference for risky alternatives can be represented and with which methods the decision maker's preference can be determined.

How a decision maker is supposed to assess risky alternatives in a normative way has been discussed in economics for quite a long time. Since an alternative a is defined by the consequences a_i and their probabilities p_i, the evaluation of an alternative has to be based on these factors in some way. In doing so, the decision maker's attitude towards the risk of the lottery and toward the value of the consequences needs to be taken into account. The objective of our further considerations is to develop a calculus of preferences that computes a value for each alternative preferably in an easy but "reasonable" way. This value is supposed to be fully analogous to value obtained for decisions under certainty and represents the decision maker's preference regarding risky alternatives.

To simplify, let us initially assume, that the consequences are being measured on an interval scale. Being a statistically trained reader, you could then suggest to define the expected value (EV) of a risky alternative a as the foundation of a rational preference. That is:

$$\mathrm{EV}(a) = \sum_{i=1}^{n} p_i \cdot a_i. \tag{9.1}$$

According to this calculus an alternative a would be preferred to an alternative b if and only if the expected value of a is greater than the expected value of b.

In the 18th century Bernoulli (1738, 1954) showed, by aid of the St. Petersburg Paradox, that decision makers deviate systematically from the preferences suggested by the expected value criterion. The St. Petersburg Lottery, upon which the paradox of the same name is based, is displayed in Figure 9-2, adjusted to today's currency.

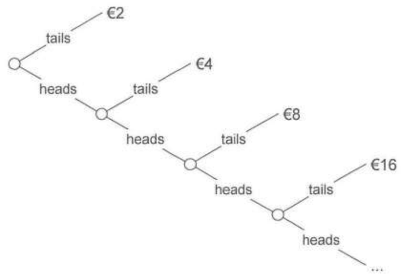


Figure 9-2: The St. Petersburg Lottery

In the St. Petersburg Lottery, the decision maker has to evaluate a multi-stage lottery. In the first stage, a fair coin is tossed. If tails comes up, the player receives €2, if heads comes up the coin is tossed once again. If the coin is tossed the second time and tails then comes up, the player receives 2^2 = €4. If not, the coin will be tossed a third time, and so forth. In the nth turn, the player receives €2^n for tails and for heads the opportunity to toss the coin once more. Since it is a fair coin, there is a 50% probability for the event heads just as for the event tails. It holds:

$$\begin{aligned}\text{EV(St. Petersburg Lottery)} &= \frac{1}{2}\cdot 2+\frac{1}{4}\cdot 4+\frac{1}{8}\cdot 8+\ldots\\ &= \sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^i \cdot 2^i \qquad (9.2)\\ &= \text{infinite.}\end{aligned}$$

The expected value of the St. Petersburg Lottery is infinite. If you were facing the decision to either participate in this lottery or to receive a certain €1 million, according to the criterion of expected value you would have to choose to participate in the St. Petersburg Lottery. However, you would most likely (like most other people) prefer the safe payment of €1 million. If one asks students – for instance within one of our decision analysis lectures – there are usually only very few that would be willing to pay more than €20 to participate in the St. Petersburg Lottery.

The lottery shows that the preference determined by the theory of expected utility does not correspond to intuitional decision behavior. A discrepancy between normatively desirable and actually exhibited behavior, however, is at first nothing out of the ordinary in the field of decision theory. The criterion of expected value, however, dictates a behavior that, even after careful consideration, is not desired by most decision makers. The St. Petersburg Lottery brings up a painful subject

for the calculus of EV: it points out implications of the calculus that are not considered rational by most decision makers. Elementary decision problems that point out a calculus' weaknesses so clearly are also termed "paradoxes". The St. Petersburg paradox hence shows the necessity of developing a better decision calculus for decisions under risk. For modeling preferences under risk, we now recommend alternative decision principles. Following Bernoulli, one could transform the payments by means of a particular value function (Bernoulli suggests a logarithmical value function) and calculate the EV of the transformed alternatives. Instead, one could also subtract a quantity from an alternative's EV that reflects the risk of the lottery. Many other evaluation rules can be derived from cogitation and have (unfortunately) found their way into the literature.

We do not want to follow this approach. Analogous to the procedure in Chapter 5 (Decisions under certainty with one objective), we want to impose requirements on a rational preference; these axioms define what we perceive as rational behavior for decisions under risk. The correct decision calculus will be deduced from these requirements on a rational preference. At the same time a measuring specification to represent the preference can be obtained.

9.2 The theory of expected utility

9.2.1 Expected utility

The theory of expected utility, which will be presented below, was established by von Neumann and Morgenstern (1947); they defined axioms and from them derived a preference calculus. After von Neumann and Morgenstern had introduced the utility theory, a whole series of axiomatic systems were developed, all of which (except for slight technical differences, cf. Fishburn 1970) led to the same calculus of expected utility. Most of the axiomatic systems are variations of the system set up by Herstein and Milnor (1953), on which we would like to base our discussion.

If the decision maker's preference regarding risky alternatives fulfills the axioms:

- complete ordering
- continuity
- independence,

then there exists a function u – called the "utility function" – whose expected value ("expected utility") represents the preference. Before we explain the axioms in more detail, the utility function will first be introduced. A lottery a will be preferred to a lottery b if and only if the expected utility of a, EU(a), is greater than the expected utility obtained for lottery b. The expected utility of a risky alternative a is defined by:

$$\text{EU}(a) = \sum_{i=1}^{n} p_i \cdot u(a_i). \qquad (9.3)$$

As discussed previously, the utility function u assigns a real number to each consequence a. The utility function u, however, represents the attitude towards the value of the consequence as well as the behavior under risk. Note the terminology to distinguish these two cases: while in the case of decision making under certainty we talk about a *value function* v, the preference for decisions under risk will be represented by a *utility function* u. This utility function is unique up to a positive linear transformation, i.e. every function u' with $u' = \alpha \cdot u + \beta (\alpha > 0)$ results in the same ordering of lotteries as with the original function u. A utility function requires no special way of scaling the consequences. You can – as was the case with the value functions – determine the value of money, of kindergarten spaces or the color of a tie.

Imagine you have asked the municipal council for 20 additional kindergarten places. The municipal council decides to create 10 new places; you are dissatisfied with this decision and are considering whether or not you should take legal action. You estimate that you would win a lawsuit with a probability of $p = 0.6$ and get 20 places and lose the lawsuit with a probability of $p = 0.4$ and get no new places at all (the municipal council would be annoyed and thus revoke its original resolution).

The expected utility criterion now gives you a clear instruction on rational decision making. You first determine the utility function by one of the methods that will be introduced later and obtain for instance: u(0 places) = 0, u(10 places) = 0,55 and u(20 places) = 1. The expected utility of the safe alternative is 0.55, the expected utility of going to court is 0.4·u(0 places) + 0.6 u(20 places) = 0.6. Being a rational decision maker, you file a suit against the ruling because the expected utility of this alternative is the largest.

Think about the following problem that you are facing under the Christmas tree while making your choice of color: you have the opportunity to give the pink tie back to your grandmother and might be able to buy one in your favorite color instead, or you could annoy her by returning the gift and end up with no tie at all. The expected utility approach can help you here as well: you have to compare the expected utility of the safe alternative (pink tie) with the expected utility of the risky alternative (favorite color versus no tie).

The criterion of expected utility gains its importance by the fact that it is not a random criterion; if a decision maker accepts the axioms that were mentioned above as a foundation of his (rational) behavior, he *must* act according to the theory of expected utility in risky decision situations: an alternative is *optimal* if and only if it has the highest expected utility of all considered alternatives.

If a decision maker rejects the axioms, or if he wants his decision behavior to be based on an alternative axiomatic system, other preference theories could possibly be deduced. We will refer to this point several times in the following and particularly go into more detail in Chapter 13. At this point, let us clearly point out that we are not aware of a theory that establishes a more convincing foundation for rational decision making than does expected utility theory.

We have not yet discussed the actual meaning of the term "utility", i.e. what conceptual meaning can be related to the word "utility" – if any at all. This question has occupied science intensely for decades. The approach followed by von Neumann and Morgenstern (and also Savage) is characterized especially by the fact that it allows rational decision making under risk without a conceptual definition of the term utility. For the deduction of the theory of expected utility, we only assume that the decision maker's preference regarding risky alternatives is observable; we do *not* need an interpretation of the utility of a consequence. The utility function is solely a function that serves to rank lotteries in a certain order – no more and no less.

9.2.2 Axiomatic foundation of utility theory

Since axioms are the central cornerstones of the theory of expected utility, we wish to explain them in more detail in the following.

Complete Ordering

Completeness: for every pair of lotteries a and b, it holds: $a \succeq b$ or $b \succeq a$. Transitivity: for all lotteries a, b, and c, it holds: from $a \succeq b$ and $b \succeq c$ follows $a \succeq c$.

The axiom of complete ordering states that arbitrary lotteries can be compared to each other and that the preference order concerning the lotteries is transitive. In principle, you already know this axiom from the chapter about value functions; here, the axiom will be defined for risky alternatives instead of safe alternatives. The axiom of complete ordering summarizes the axioms of completeness and transitivity that were both introduced in Chapter 5.

In this chapter, it should again be discussed whether or not the completeness claim is really reasonable. One will face situations, in which a decision maker cannot make a preference statement or an indifference statement regarding two lotteries. For example, who would like to compare a 1% chance of dying in a car accident with a 1% chance of dying in a bicycle accident? From a prescriptive point of view, a decision maker should have a preference in this sad case as well; the theory of expected utility should be able to make preference statements for any pair of alternatives and completeness is an indispensable precondition for it. The transitivity claim could be disputed as in Chapter 5 as well. However, we claim transitivity as the foundation of rational behavior.

Continuity

Given lotteries a, b, c with $a \succeq b \succeq c$, there exists some probability p, for which $b \sim p \cdot a + (1-p) \cdot c$.[1]

The term $p \cdot a + (1-p) \cdot c$ describes a compound lottery – a lottery whose results are again lotteries – for which lottery a occurs with a probability p and lottery c occurs with a probability $1-p$. For example, if a = (€100, 0.8; €0, 0.2) and c = (€200, 1), then the lottery $0.5 \cdot a + 0.5 \cdot c$ equals (€200, 0.5; €100, 0.4; €0, 0.1).

---

[1]In order to simplify we assume that all utility functions are bounded and the expected utilities of all lotteries are hence finite.

Figure 9-3 displays the lottery $0.5 \cdot a + 0.5 \cdot c$ as a two-stage and equivalent one-stage lottery.

The axiom of continuity implies that for each lottery b that lies between some lotteries a and c in terms of preference, a combination of a and c that is as good as b can always be found. In many cases, the axiom of continuity is considered as a condition that one needs for the mathematical derivation of the theory of expected utility but which does not have to be discussed further; in particular cases, however, this axiom's validity will nevertheless have to be discussed. If lottery a corresponds to a sure gain of €2, lottery b corresponds to a sure gain of €1, and lottery c represents a severe car accident during the next week, many decision makers will not be able or not be willing to name a combination of a and c such that they are indifferent between this lottery and lottery b. Even the highest probability p (which, however, has to stay smaller than one) does not lead to any indifference statement. If one refrains from such admittedly constructed examples, the axiom of continuity is undisputed and a generally accepted foundation of rational behavior.

The axioms of continuity and complete ordering are basically identical to the requirements on preferences that guarantee the existence of a value function for decision making under certainty. Like in that case, the existence of a preference function for risky alternatives, that orders lotteries according to the decision maker's preference, can be derived from these two axioms. The statement that a preference function exists, however, is not of much help when it comes to finding a solution for the decision problem. The distinctive feature of the theory of expected utility is its representation of the decision maker's preference as a simple function of probabilities and consequences. The axiom that gives von Neumann and Morgenstern's calculus for decision making under risk its special form and that is hence responsible for the simple, additive model of expected utility theory is the axiom of independence.

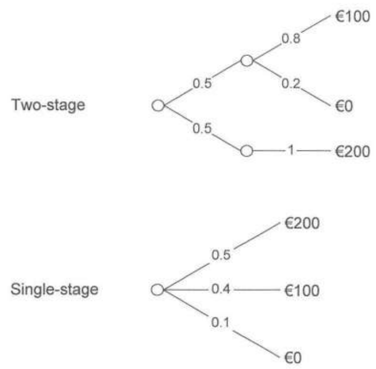


Figure 9-3: Compound lottery: Two-stage and equivalent single-stage lottery

Independence

If $a \succeq b$ holds for two lotteries, then it also has to hold for all lotteries c and all probabilities p, that $p \cdot a + (1-p) \cdot c \succeq p \cdot b + (1-p) \cdot c$.

The independence axiom for decisions under risk is based on the same idea outlined in Chapter 6 concerning the independence conditions for decisions under multiple objectives: a preference between two lotteries a and b should not change if both lotteries are connected to the same lottery that is hence irrelevant to this decision problem.

We want to explain the independence axiom by means of an example (cf. Figure 9-4). A decision maker prefers lottery $a = (€100, 0.5; €0, 0.5)$ to lottery $b = (€60, 0.7; €10, 0.3)$. This preference must remain, if he mixes both lotteries a and b with lottery $c = (€50, 1)$ for all probabilities. In Figure 9-4, we choose $p = 0.8$. The mixture can be represented as a two-stage or as a single-stage lottery. In the two-stage representation, one can see the implications of the independence axiom clearly and precisely: the 20% probability of obtaining c is combined with a 80% probability of obtaining lottery a or lottery b.

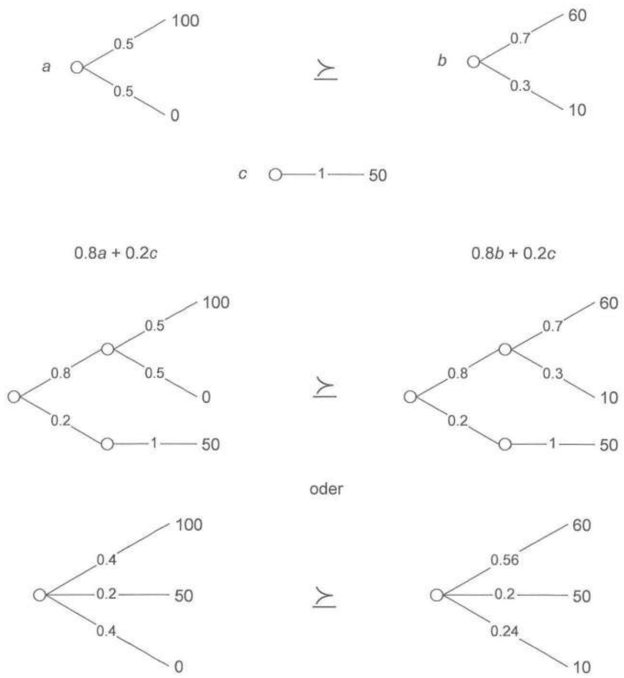


Figure 9-4: Example for the independence axiom of utility theory

It is easy to see that a decision maker's indifference between lotteries a and b implies indifference between the combined lotteries $p \cdot a+(1-p) \cdot c$ and $p \cdot b+(1-p) \cdot c$ (as indifference implies $\succeq$ as well as $\preceq$); this is why the independence axiom is often called the substitution axiom. A lottery (or a consequence) may be substituted by another lottery if the decision maker is indifferent between both lotteries or respectively between the lottery and the consequence. The substitution has no impact on the decision maker's preference. This statement can be clarified by means of Figure 9-5. Alternative b was derived from alternative a by replacing the consequence €100 in alternative a by the lottery (€250, 0.5; €0, 0.5). If the decision maker is indifferent between the consequence €100 and the lottery (€250, 0.5; €0, 0.5), he also has to be indifferent between alternative a and alternative b.

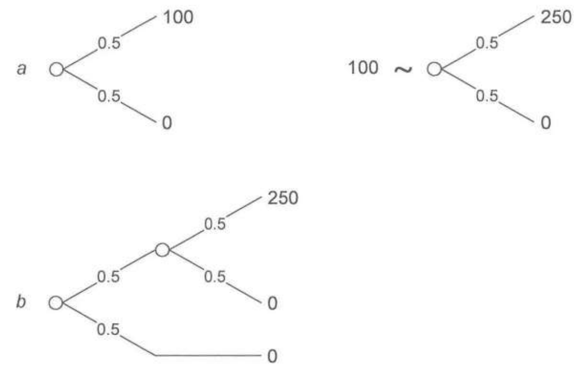


Figure 9-5: Example for the substitution axiom

In our discussion, we have so far implicitly assumed the validity of yet another axiom. The axiom of the reduction of compound lotteries states that a decision maker must be indifferent between a two-stage lottery (middle pair comparison in Figure 9-4) and a one-stage lottery whose results are equivalent to the two-stage lottery's results and whose probabilities were derived by multiplication of the corresponding two-stage lottery's probabilities (lower pair comparison in Figure 9-4). This axiom touches much more elementary ground. It belongs to the class of axioms that postulate invariance under the different description of a decision problem, i.e. that the description or representation of a decision problem should not affect a decision maker's choice as long as the relevant information (the consequences and their associated probabilities) is the same. In fact, as we will see in Chapter 13, this claim is by no means warranted when ad hoc decisions are made. Therefore it is important to mention this axiom, even though it works on quite a different level. Clearly, a violation of such an elementary axiom allowing the multiplication of probabilities for compound lotteries immediately implies a violation of Kolmogoroff's axioms of probability calculus (cf. Chapter 7); it should thus be one of the building blocks of rational decision making.

The independence axiom significantly limits the amount of acceptable preferences. Given a preference for two alternatives a, b, it implies a preference for all compositions with arbitrary alternatives c. In other words: the drawback of the simple shape of expected utility's preference function is a strong restriction concerning the underlying preference. As we will show for the representation of more recent descriptive theories in Chapter 13, the independence axiom is violated intuitively and cannot serve to describe decision behavior. This violation of the axiom when making intuitive decisions has even led to the rejection of expected utility theory as a foundation for rational behavior (Allais 1953). Meanwhile the separation between prescriptive and descriptive perception, made by us as well, has hardened in literature. Rational decision making shall exactly be characterized by the facts that preferences that are independent of shared components in alternatives, that it is possible to substitute equally preferred components without affecting the overall decision, and that the commonly accepted multiplication of probabilities for compound lotteries is allowed. Therefore, the independence axiom is affirmed in most of the literature from a prescriptive perspective, which we will follow here. This discussion also points out that it is only an examination of the axioms a theory is based on that allows us to reasonably discuss whether a theory can be the foundation for "rational decision making".

Axioms and expected utility formula

Presumably, you are wondering why these three mentioned axioms (complete ordering, continuity, and independence) imply expected utility's evaluation rule as in equation (9.3). Fortunately, the argumentation is not overly complex and we shall briefly sketch it here (for a more general overview, cf. Herstein and Milnor 1953).The basic idea is that the three axioms allow any arbitrary lottery to be represented as a lottery that is equally preferred and consists only of some combination of the outcomes x_{max} and x_{min} making it easy to compare lotteries. x_{max} and x_{min} here denote the best and worst consequences possible within the context of the decision problem. Take a look at alternative a in Figure 9-6; according to the continuity axiom there is a lottery $(x_{max}, q_i; x_{min}, 1-q_i)$ for each consequence a_i, so that the decision maker is indifferent between a sure amount a_i and this lottery. According to the independence axiom the decision maker is moreover indifferent between lottery a and lottery a' (the consequences may be substituted by equivalent lotteries). A reduction of the two-stage lottery to its one-stage equivalent leads to lottery a'' which is thus equally preferred to lotteries a and a'.

The transformation from a to a'' can be conducted similarly for every other arbitrary lottery b. Owing to the indifference between a and a'' as well as between b and b'', the preference concerning a and b corresponds to the preference concerning a'' and b''. This latter preference, however, is easy to determine: the strength of preference for lottery a (or b) is expressed in the probability of receiving x_{max} in lottery a'' (or b''). A lottery a should hence be preferred to a lottery b if the probability of receiving x_{max} is greater for lottery a'' (which was transformed from a) than for lottery b'' (which was transformed from b). The probability of winning x_{max} thus induces an ordering of the lotteries in line with the decision maker's preference. If one now defines $u(a_i) = q_i$, it becomes apparent that the probability of achieving x_{max} equals the alternative's expected utility. Preferences of a rational decision maker are hence based on the expected utility of the alternatives.

9.2.3 The three-outcome-diagram

The implications of the independence axiom and hence the special properties of utility theory can be represented in so-called *three-outcome-diagrams* with the aid of utility indifference curves (Machina 1982, Schauenberg 1990, and Weber and Camerer 1987). Let x_l, x_m, x_h (low, medium, high) be three arbitrary consequences

for which it holds true that $x_l \prec x_m \prec x_h$. All lotteries with these consequences can be represented in the same two-dimensional three-outcome-diagram. The probabilities p_l and p_h, that lead to the consequences x_l and x_h are plotted on the axes of this diagram (cf. Figure 9-7). Since it holds that $p_m = 1 - p_l - p_h$, the whole lottery is characterized by the two probabilities p_l and p_h. In Figure 9-7, the lottery $(x_h, 0.62; x_m, 0.26; x_l, 0.12)$ is exemplarily marked as point G.

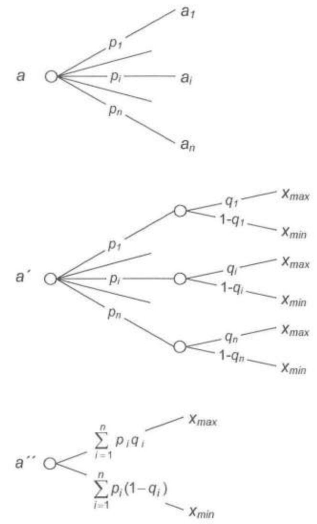


Figure 9-6: Derivation of expected utility formula from the axioms

A decision maker's preference can be clarified by utility indifference curves in a three-outcome-diagram. Let Eu^* be a certain utility level that can be achieved by a lottery with the three consequences x_l, x_m, x_h. If one writes down the equation for the expected utility of these lotteries $(x_h, p_h; x_m, p_m; x_l, p_l)$ and $(\text{EU} = u(x_h) \cdot p_h + u(x_m) \cdot p_m + u(x_l) \cdot p_n)$, sets them equal to Eu^* and solves them for p_h (note that $p_m = 1 - p_l - p_h$), one obtains:

$$p_h = p_l \frac{u(x_m) - u(x_l)}{u(x_h) - u(x_m)} + \frac{Eu^* - u(x_m)}{u(x_h) - u(x_m)}. \tag{9.4}$$

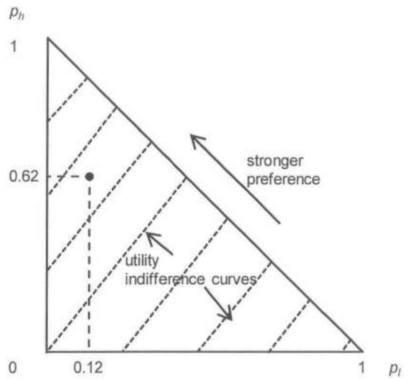


Figure 9-7: Three-outcome-diagram

For every Eu^*,p_h is a linear function of p_l with identical slope, implying that the indifference curves are parallel straight lines as shown in Figure 9-7. An increase of the utility level Eu^* causes a parallel shift of the utility indifference curves to the upper left.

We will next discuss the connection between the independence axiom and the parallelism of the indifference curves again for a concrete example (which we will encounter again in Chapter 13 under the name Allais paradox). Consider the lotteries a=(€3,000, 1) and b=(€4,000, 0,8; €0, 0,2). They can be marked in a three-outcome-diagram with the consequences x_h=€4,000, x_m=€3,000, and x_l=€0 as in Figure 9-8.

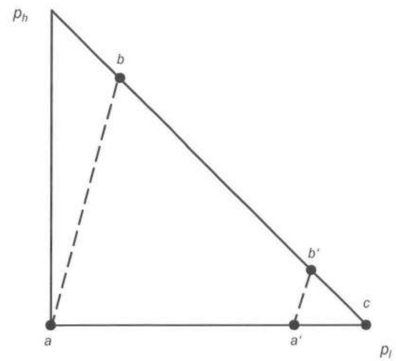


Figure 9-8: Lotteries of the Allais paradox

We now combine these two lotteries with the lottery c = (€0, 1). Under the assumption that lottery c occurs with a probability of 75%, we obtain the lotteries $a' = 0.25 \cdot a + 0.75 \cdot c$ = (€3,000, 0.25; €0, 0.75) and $b' = 0.25 \cdot b + 0.75 \cdot c$ = (€4,000, 0.2; €0, 0.8). These two lotteries are marked in the three-outcome-diagram in Figure 9-8 as well. Interestingly, the line segment between a and b and the line segment between a' and b' have the same slope (and this is not specific to this example, but is generally the case). Thus, the parallel straight lines representing the indifference curves explain the validity of the independence axiom. The indifference curves of a decision maker are either steeper than the line segment between a and b (implying $a \succ b$) as in the left panel of Figure 9-9 or they are flatter (which implies $b \succ a$) as in the right panel of Figure 9-9. Anyway, the preference translates to the preference over the lotteries a' and b' owing to the parallelism of the indifference curves. An expected utility maximizer cannot simultaneously have the preferences $a \succ b$ and $a' \prec b'$.

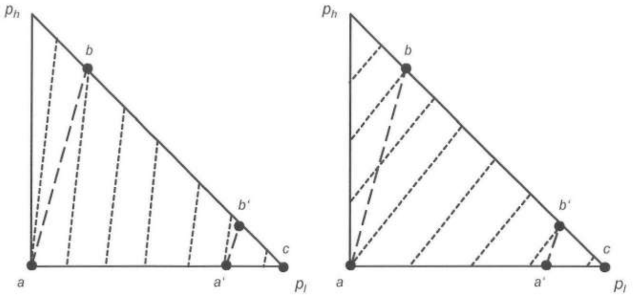


Figure 9-9: Possible indifference curves for the Allais paradox

The three-outcome-diagram shows that expected utility theory can only display a very restricted amount of preferences although it imposes no restriction on the utility function u: many other forms other than parallel straight lines would also be generally conceivable. Fixing one indifference curve determines the preference order for all lotteries that can be displayed in the diagram. This property results directly from the formula of expected utility that forms the basis of the upper equation of the indifference curve. If a decision maker would decide according to the expected *value* criterion, the indifference curve's slope would also be given. The exact slope in this case, however, is determined by the relative location of x_h, x_m and x_l. Generally, steeper indifference curves in a three-outcome-diagram (meaning that x_h, x_m and x_l are fixed) represent higher risk aversion. To see this, consider the two indifference statements:

$$x_m \sim (x_h, 50\%; x_l, 50\%)$$
$$x_m \sim (x_h, 70\%; x_l, 30\%)$$

The lottery in the second statement has a higher expected value, but the same certainty equivalent (CE) (namely x_m) as in the first statement. The risk premium is hence higher in the second statement. The indifference curves for both statements are straight lines through the origin, which represents the consequence x_m. The first line passes through the point (50%, 50%) whereas the second line passes through (30%, 70%). The second line representing the second indifference statement is hence steeper.

Recent theories for the description of decision behavior (cf. Chapter 13) would like to allow for a wider spectrum of preferences compared to expected utility theory and hence be able to predict "paradox" behavior as well. Consequently, these theories have indifference curves in the three-outcome-diagram that are either not parallel or not even linear, these theories can model preferences with indifference curves in Figure 9-8 such that $a \succ b$ and $a' \prec b'$.

9.2.4 Subjective expected utility theory

The expected utility theory as developed by von Neumann and Morgenstern, does not deal with the problem of how to determine the probabilities needed for decision making. It simply assumes that the probabilities are given or were determined independently of the calculation of the utility function. This assumption, however, is not fulfilled in many real-world situations. In Chapter 7, the methods of determining probabilities are described. In the indirect methods, the decision maker is presented with two lotteries for which he has to alter one of the lottery's probabilities (symbolized for example by a decision wheel) so that he is indifferent between both lotteries; the subjective probability of the event that had to be rated was derived from this indifference statement. The idea that probability statements can be derived from preference statements was converted into an axiom by Savage (1954, 1972) and is the most accepted foundation of rational decision making under risk. Many new descriptive theories expand the approach developed by Savage.

In Savage's approach, a decision maker chooses between alternatives a (*acts*) that depend on *states of nature* $s \in S$ (also called events) and lead to *consequences* $a(s)$. Preferences regarding alternatives allow conclusions about utility functions and "personal plausibilities" for the occurrence of events. Under certain axioms for the decision maker's preference, Savage was able to show that the personal plausibilities satisfy the claims on probabilities; they are hence called decision weights or subjective probabilities.

Savage's theory is called *subjective expected utility theory* (SEU). It evaluates the alternatives completely analogous to utility theory with the sum of the products of probability and utility. For the subjective expected utility of an alternative a, it holds then:

$$\mathrm{SEU}(a) = \sum_{s \in S} p(s) \cdot u(a(s)). \tag{9.5}$$

At this point, we could present all of Savage's axioms and hence show in particular which claims have to be made on preference statements to be able to derive subjective probabilities. However, we only want to present Savage's most important axiom here. Since Savage derives the expected utility as the sum of the products of subjective probability and the consequence's utility, he needs an independence axiom as well which is called the "Sure thing principle". The independence axiom of SEU of course resembles the independence axiom of utility theory. Also of interest is the connection to the independence axioms for decisions with multiple objectives that were presented in Chapter 6. An additive representation of the preference – be it additive over states or objectives – always requires an independence axiom's validity (cp. generally on independence conditions for information aggregation Dyckhoff 1986).

Independence for SEU

Let a, b, a' and b' be alternatives and let S' be a subset of the set of events S with $a(s) = a'(s)$ as well as $b(s) = b'(s)$ for $s \in S'$ and $a(s) = b(s)$ as well as $a'(s) = b'(s)$ for $s \in S \setminus S'$, then $a \succ b$ holds if and only if $a' \succ b'$.

Put into words, this independence axiom states: if two alternatives have identical consequences for certain events then these events cannot influence the decision maker's preference regarding these alternatives. If one replaces events by objectives, the connection to the axioms for multiple objectives becomes clear. The independence axiom can be explained with an example: a die is tossed and according to Table 9-1, you receive either an apple (A) or a banana (B) depending on the number thrown.

With lottery a one gets a banana independent of the number thrown, for b one obtains a banana if 1 till 5 come up and an apple if 6 comes up; for b' one always gets an apple and for a' one only gets a banana if 6 comes up and an apple otherwise. The independence axiom states that the preference between a and b has to be identical to the preference between a' and b'. The consequences of states 1 till 5, which are identical for both alternative pairs, cannot play a decisive role for the decision. In state 6, a comparison of a and b and a comparison of a' and b' must lead to the same preference.

Table 9-1: Explanation of the "Sure thing principle"

	1	2	3	4	5	6
A	*B*	*B*	*B*	*B*	*B*	*B*
B	*B*	*B*	*B*	*B*	*B*	*A*
a'	*A*	*A*	*A*	*A*	*A*	*B*
b'	*A*	*A*	*A*	*A*	*A*	*A*

Now that you have learned the main aspects of utility theory, you could ask, how the utility function (which appears to be essential for one's preferences) can be determined. We will learn about methods of determining the utility function in Section 9.4. Before this, we want to define fundamental terms of utility theory in Section 9.3.

9.3 Basic concepts of utility theory

9.3.1 Certainty equivalent

A central term for the evaluation of lotteries is the certainty equivalent (CE). Looking at a lottery a, the certainty equivalent CE(a) represents the safe consequence for which the decision maker is indifferent between CE(a) and the lottery that has to be evaluated. It hence holds that

$$u(\text{CE}(a)) = \text{EU}(a). \tag{9.6}$$

In case of a continuous set of consequences and a continuous utility function, the decision maker has to be able to name a certainty equivalent according to the axiom of complete preference ordering; in other cases, this is not necessarily true (Laux 2007, pp. 221-222.).

9.3.2 The risk attitude

Look at the three strictly monotonically increasing utility functions u_{RN}, u_{RP} and u_{RA} displayed in Figure 9-10 that could be defined over any continuous consequences (money, amount of rain, amount of tomato sauce). The minimal consequence is denoted as x_{min}, the maximal consequence is denoted as x_{max}. The functions differ with respect to their curvature; the following allocation holds, and the seemingly cryptic notation will soon become clear:

u_{RN} linear utility function
u_{RP} convex utility function
u_{RA} concave utility function.

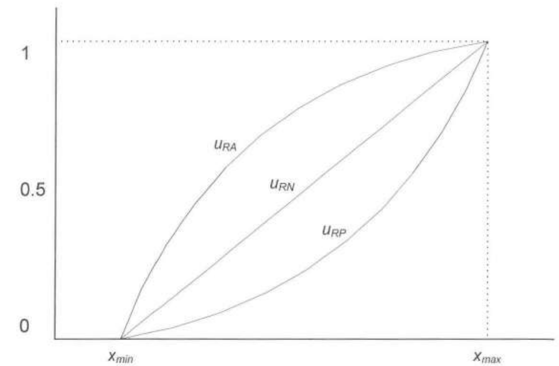


Figure 9-10: Linear, convex and concave utility functions

It now suggests itself to examine if the curvature (i.e. the shape) of the utility function allows for general statements about decision behavior under risk. Let us hence look at an arbitrary lottery a. According to the certainty equivalent's definition, the following relation holds:

$$u(\text{CE}(a)) = \text{EU}(a), \text{ i.e. } \text{CE}(a) = u^{-1}(\text{EU}(a)). \tag{9.7}$$

These equations show how each lottery's certainty equivalent can be derived via the utility function.

In order to obtain an insight into decision behavior, the certainty equivalent will now be contrasted with the lottery's expected value. The difference of these two values is called the risk premium (RP):

$$\mathrm{RP}(a) = \mathrm{EV}(a) - \mathrm{CE}(a). \tag{9.8}$$

Given for instance a lottery (€100, 0.5; €0, 0.5) for which the decision maker claims his certainty equivalent to be €40, the risk premium would hence be €10 for this lottery. The risk premium serves to characterize the risk attitude. Note that CE(a) and RP(a) may well vary among different decision makers whereas EV(a) is a characteristic of the lottery and thus independent of the decision maker. For monotonically *increasing* utility functions, the following holds: if RP(a) > 0 for all risky lotteries a, one refers to the decision maker as risk averse, if RP(a) = 0, one refers to him as risk neutral and if RP(a) < 0 one refers to him (and hence to his behavior) as risk prone. If a decision maker is risk averse, his risk premium represents the amount which he – based on the lottery's expected value – is willing to forego to avoid the risk of the lottery and instead receive the certainty equivalent for sure.

In a last step, we want to look for a connection between the curvature of the utility function and the risk attitude. The connections can be seen in the following table; that shows that the curvature of the utility function determines the risk attitude. If we look at the concave function u_{RA} from Figure 9-10, for example, we now discover that it displays risk averse decision behavior, i.e. given this type of utility function, the certainty equivalent is always smaller than the expected value for every arbitrary lottery.

Table 9-2: Risk attitude and curvature of the utility function

Utility function	RP(=EV−CE)	Risk attitude
linear: u_{RN}	=0	risk neutral
concave: u_{RA}	>0	risk averse
convex: u_{RP}	<0	risk prone

We want to give an example to illustrate the connection between the curvature of the utility function and the decision maker's risk attitude. Assume the decision maker has a concave utility function over the interval $[x^-, x^+]$ and the two-state lottery $a = (a_1, p; a_2, 1-p)$ has consequences in this interval. Figure 9-11 depicts this situation. The expected utility EU(a) is marked on the y-axis, the expected value EV(a) on the x-axix. In this example with only two consequences, the point (EV(a), EU(a)) lies on the line segment between the two points on the utility function that represent the two consequences (the exact position on the line segment depends on the probability p; the point (EV(a), EU(a)) splits the line according to the ratio $(1-p)$ to p). The intuition for the certainty equivalent and the risk pre-

mium are as follows: the horizontal line that passes through EU(a) intersects with the utility function. Project this intersection point onto the x-axis. We have then found the consequence x that has the same utility as the lottery a. According to our definition, this x is the certainty equivalent of a. As becomes clear now, for a concave utility function, the certainty equivalent is always below the expected value EV(a). On the x-axis, the distance between the expected value and the certainty equivalent corresponds to the risk premium. This example nicely illustrates that a more concave utility function would result in a higher risk premium.

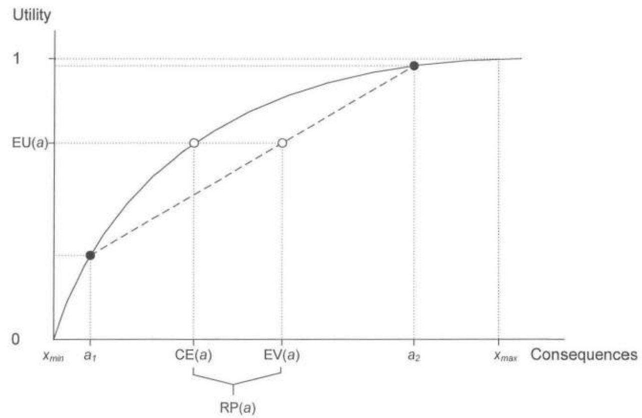


Figure 9-11: Concave utility function, certainty equivalent, and risk premium

A decision maker's risk attitude is also reflected in the indifference curves in the three-outcome-diagram. As shown previously in Figure 9-7 and the equation of the utility indifference curves, a stronger risk aversion is reflected in a "more concave" utility function. Keeping $u(x_l)$ and $u(x_h)$ constant, the increase of risk aversion is hence reflected in an increase of the term $u(x_m)$. This increase of $u(x_m)$ leads to a growth of the coefficient of p_l: the indifference curves in the diagram become steeper.

So far, we have limited our examination to increasing utility functions. For monotonically *decreasing* utility functions (think about a utility function on the number of days that you have to stay in hospital, for example) one speaks of risk-averse behavior if RP < 0 and if the utility function is concave; conversely, one speaks of risk-prone behavior if RP > 0 and if the utility function is convex. If the set of consequences is not continuous or if the consequences are ordinally or nominally scaled, the definitions above cannot be used.

Concerning identical decision problems, the risk attitude as expressed by the utility function can vary among different decision makers. Even for the same decision maker, different utility functions and hence different risk attitudes may arise in different decision situations – at least in a broad sense. When taking out insurance policies, one is usually risk averse (as one prefers the safe payment for the insurance premium to the risky damage whose expected loss is smaller than the risk premium); if one plays the lottery, one is usually risk prone (one prefers the risky lottery whose expected benefit is smaller than the safe payment for a lottery ticket). Therefore, the risk attitude has to be re-determined for different decision situations (in a broad sense) and for every decision maker. Some areas of business administration, however, take general assumptions about the risk attitude of decision makers as a basis.

9.3.3 The Arrow/Pratt measure of risk aversion

So far we have learned that we can classify decision makers into three categories (risk averse, risk neutral and risk prone) according to their risk attitude. Now, we want to define an exact measure that can also quantify the risk attitude. With this, we will be able to classify decision makers as more or less risk averse – and so on. The risk attitude is reflected in the strength and kind of curvature of the utility function. This curvature may be measured by the Arrow/Pratt measure of risk aversion $r(x)$:

$$r(x) = -\frac{u''(x)}{u'(x)}. \tag{9.9}$$

The measure requires that the utility function is twice differentiable and that the first derivative of the utility function is different from zero. In the form given above, the measure gauges the decision maker's absolute risk attitude. The measure is absolute, as it will not be put in relation to the lottery's consequences for which it was computed. By multiplication with the respective consequence x, one obtains a measure for the relative risk attitude $r^*(x)$ (also referred to as proportional risk attitude):

$$r^*(x) = -\frac{u''(x)}{u'(x)} \cdot x. \tag{9.10}$$

Both quantities are needed in finance theory to characterize the risk behavior of investors. One usually assumes (Kraus and Litzenberger 1976) that:

- money's marginal utility is positive;
- the marginal utility decreases with increasing amounts of money; and
- absolute risk aversion is not increasing.

As has already been argued, risk premium and curvature of the utility function – and hence risk premium and risk attitude measures – depend directly on each other:

$$\mathrm{RP} \approx ½ \cdot \mathrm{var(Lotterie)} \cdot r(\mathrm{CE}). \tag{9.11}$$

The risk premium is approximately equal to the product of half of the lottery's variance and the risk attitude measure at the position of the certainty equivalent. A detailed derivation can be found in Pratt (1964), who also shows how the risk premium can be approximated even more precisely by higher moments of the lottery. The formula above shows an approximate connection between three important quantities, showing that a greater (smaller) variance of an alternative leads to a greater (smaller) risk premium.

You can immediately see from this formula that risk premium and risk attitude measures always have the same sign. For monotonically increasing functions, it holds that if $r(x) > 0$ for all x (that is $\mathrm{RP} > 0$) then the decision maker is risk averse; if $r(x) < 0$ for all x (that is $\mathrm{RP} < 0$) then the decision maker is risk prone; for the case $r(x) = 0$ (that is $\mathrm{RP} = 0$) risk neutrality holds and the utility function is linear. It analogously holds for monotonically decreasing utility functions that $r(x) < 0 \Leftrightarrow \mathrm{RP} < 0 \Leftrightarrow$ risk aversion and $r(x) > 0 \Leftrightarrow \mathrm{RP} > 0 \Leftrightarrow$ risk proclivity.

The question about the absolute and relative risk attitude of "normal" individuals has intensely occupied the research. For example, see Friend and Blume (1975) who derived on the basis of American tax data that the average investor has decreasing absolute and constant relative risk aversion.

9.3.4 Risk attitudes of selected utility functions

In this section, we want to introduce some important utility functions and characterize them with respect to the risk attitude connected with them. This can facilitate the work concerning practical determination of utility functions and provide helpful service within economic theory construction.

- The exponential utility function is:

$$u(x) = \alpha + \beta e^{-cx} \quad \text{with} c > 0 \text{ and } \beta < 0. \tag{9.12}$$

It exhibits constant absolute risk aversion and increasing relative risk aversion.

- The quadratic utility function is:

$$u(x) = \alpha + \beta x - \gamma x^2 \quad \text{with} \beta, \gamma > 0, \;\; x \le \frac{\beta}{2\gamma}. \tag{9.13}$$

This function exhibits increasing absolute risk aversion as well as increasing relative risk aversion. This behavior will normally be observed neither in experimental studies nor in real life. The quadratic utility function should thus only be used with utmost caution (best choice: not at all) to represent human behavior.

- The logarithmic utility function is:

$$u(x) = \alpha + \beta \log(x) \quad \text{with} \beta > 0. \tag{9.14}$$

It has the property of decreasing absolute risk aversion and constant relative risk aversion.

Utility functions exhibiting constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA) play a particularly important role in economics. The exponential and the logarithmic functions are representatives of these two types of risk aversion; we now take a closer look at the properties of CARA and CRRA utility functions.

For the exponential utility function in (9.12), you can easily check that $r(x)=c$, meaning it exhibits CARA (a direct consequence of this is increasing relative risk aversion as constant $r(x)$ implies that $x{\cdot}r(x)$ is monotonically increasing in x). Conversely, one can show that every utility function that exhibits CARA has this particular functional form. Therefore, no CARA functions other than the exponential functions exist.

Interestingly, decision makers with a CARA utility function choose the same absolute risks independently of their current wealth levels. Assume a decision maker has a current wealth level of W and has the opportunity to invest an absolute amount a of his wealth in a lottery L. His terminal wealth level is the sum of the remaining (i.e. not invested) wealth $W–a$ and the lottery's pay-off $a{\cdot}L$. The decision maker's goal is to maximize his expected utility by an optimal choice of a, i.e.

$$\max_a Eu(W-a+a\cdot L).$$

The term in every consequence of the decision maker's terminal wealth can be extracted and is just a scaling constant due to the well known rules of exponentiation ($\exp(W-a+a\cdot L)=\exp(W)\cdot\exp(-a+a\cdot L)$). Hence:

$$\max_a Eu(W-a+a\cdot L)=u(W)\cdot\max_a Eu(-a+a\cdot L)\,. \tag{9.15}$$

The current wealth level thus does not alter the optimal choice of the absolute amount a to be invested in the lottery.

The special functional form of the exponential utility function also guarantees that a lottery's certainty equivalent increases by δ if δ is added to all consequences of this lottery. Since in this case the expected value also increases by δ, the risk premium RP remains unchanged. This is another important property of constant absolute risk aversion: a lottery's risk premium is independent of the decision maker's current wealth.

You already know the logarithmic utility function as one example of the class of utility functions exhibiting constant relative risk aversion. It is easy to check that in this case $r^*(x)=x{\cdot}r(x)$ is a constant. The class of CRRA utility functions, however, is not limited to the logarithm. More generally, if a utility function exhibits constant relative risk aversion, it must have the form

$$u(x)=\alpha+\beta\cdot\begin{cases}\dfrac{x^{1-c}}{1-c},\ \text{if } c\neq 1\\ \log(x),\ \text{if } c=1\end{cases}\quad \text{with } \beta>0. \tag{9.16}$$

Distinguishing the two cases of $c=1$ and $c\neq 1$ is necessary to avoid a division by zero in the upper branch of the definition of u. We chose the scaling by $1/(1-c)$ because in this case slope and curvature of both branches of the definition continuously correspond to each other (in particular, the logarithm is a limiting case of CRRA utility functions as c goes to one).

The intuition of constant relative risk aversion is best explained by means of an investment example. Assume a decision maker has current wealth W and wants to invest a relative amount of a into a lottery L; for instance, he invests 60% of his current wealth. His terminal wealth is then the sum of the wealth not invested $W{\cdot}(1-a)$ and the lottery's pay-off $W{\cdot}a{\cdot}L$. Again, his goal is to maximize expected utility by an optimal choice of the fraction a, i.e.

$$\max_a Eu(W\cdot(1-a)+W\cdot a\cdot L).$$

For constant relative risk aversion, the decision maker's utility function has the form of (9.16) and the expected utility maximization problem can be written as:

$$\max_a Eu(W\cdot[(1-a)+a\cdot L])=\begin{cases}W^{1-c}\cdot\max_a Eu((1-a)+a\cdot L),\ \text{if } c\neq 1\\ \log(W)+\max_a Eu((1-a)+a\cdot L),\ \text{if } c=1\end{cases} \tag{9.17}$$

i.e. the optimal choice of the fraction a is independent of the current wealth level W. For CRRA the invested amount $a{\cdot}W$ is proportional to the decision maker's current wealth level W. In absolute terms, the invested amount increases with the wealth; this fact reflects the decreasing absolute risk aversion of CRRA utility functions.

As with CARA utility functions, we find interesting properties related to the certainty equivalent and the risk premium. If we scale all consequences of a lottery by a factor α then the certainty equivalent will be scaled by the same factor α. Obviously, the same scaling applies to the expected value of the lottery and thus to the risk premium. Relative to the risky investment the risk premium, thus, stays constant. For an even more detailed discussion of CRRA utility functions, the reader is referred to Wakker (2008).

9.4 The determination of the utility function

9.4.1 The basic-reference-lottery

In the course of this chapter we have so far essentially made theoretical deliberations on utility theory. To apply the theory of expected utility to practical problems it is indispensable to determine the decision maker's utility function regarding the respective dependent variable. If only a few consequences have to be evaluated, only the corresponding points of the utility function need to be determined. For the determination of the function, it is reasonable to remember the problems and possible solutions in the determination of value functions and probabilities. In all methods we were confronted with similar problems: the decision maker does not always have an exact preference and can be influenced in his statements by the method of elicitation. It is therefore absolutely necessary to study the methods for the determination of utility functions that will be presented in the following not only theoretically but also in practice with applicatory examples. This is the only way of learning how to get a definite utility function from initially inconsistent preferences via feedback.

In the prescriptive version of expected utility theory, the same approach as in other areas of decision theory is chosen: from the simple to the complex. The decision maker's preference, i.e. the utility function, is determined by evaluating simple, risky alternatives. If the preference is sufficiently displayed by the utility function, it can then serve to compute the optimal alternative even in complex decision situations.

The basis of most methods of determining utility functions is the so-called basic-reference-lottery (BRL) and its certainty equivalent CE*. The basic-reference lottery is displayed in Figure 9-12.

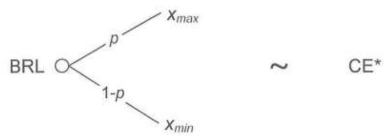


Figure 9-12: Basic-reference-lottery and certainty equivalent

At first we want to assume, that the decision maker has a monotonic utility function. The basic-reference-lottery is a lottery with the two consequences x_{max} and x_{min}, whereas the quantities x_{max} and x_{min} are the maximal and minimal possible results in a certain decision situation. The consequences occur with the probabilities p and $1-p$. The quantity CE* from the interval $[x_{min}, x_{max}]$ is the certainty equivalent of the basic-reference-lottery; the expected utility of the basic-reference-lottery thus equals the utility of CE*. We get:

$$\text{EU(BRL)} = p \cdot u(x_{max}) + (1-p) \cdot u(x_{min}) = u(\text{CE*}). \tag{9.18}$$

Since the utility function is interval-scaled, we can arbitrarily choose the points of origin and unit; we set the values $u(x_{min}) = 0$ and for $u(x_{max}) = 1$. With this we get:

$$\text{EU(BRL)} = p = u(\text{CE*}). \tag{9.19}$$

The expected utility of the basic-reference-lottery and hence the certainty equivalent's utility is equal to the probability p.

In the comparison between basic-reference-lottery and certainty equivalent, four quantities need to be considered: the probability p, both of the lottery's consequences x_{min} and x_{max} plus the certainty equivalent CE*. The methods of determining the utility function differ in which of these four quantities are given and which of these four quantities are asked for. For example, the lottery can be given and the subject is then asked for the certainty equivalent (*certainty equivalent methods*). It could also be the case that the certainty equivalent is given and he is asked for the probability (*probability equivalent methods*) or for the consequences.

The indifference between the basic reference lottery and its certainty equivalent enables us to infer the additional point (CE*, u(CE*)) on the utility function because of the three utility values ($u(x_{max})$, $u(x_{min})$ and u(CE*)) in equation (9.18), two are known and the third can be calculated with the help of the probability p. This idea of eliciting points on the utility function is not limited to the case where the lottery involved is based on the known consequences x_{max} and x_{min}. For an arbitrary lottery with consequences x_1 und x_2, as depicted in Figure 9-13, we can infer from the indifference:

$$\text{EU(RL)} = p \cdot u(x_1) + (1-p) \cdot u(x_2) = u(\text{CE*}) \tag{9.20}$$

the utility value u(CE) once we know $u(x_1)$ and $u(x_2)$. We denote such a lottery as the reference lottery. Yet another way to elicit a point on the utility function is possible: if u(CE) and one of the utility values $u(x_1)$ and $u(x_2)$ is known, the still-unknown utility value ($u(x_1)$ or $u(x_2)$) can be inferred from the indifference statement. In the following, this approach will be applied as well.

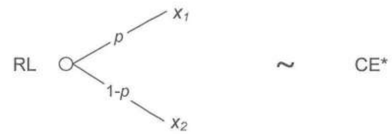


Figure 9-13: Reference lottery and certainty equivalent

Once one has identified these points, the utility function can again be completely determined analogously to the way we used for value functions; the points are interpolated linearly or a given functional form is fitted.

In the following, we want to present five simple methods of determining the utility function. For the presentation of the methods we will assume that the consequences are measured on a continuous scale. Explanatory examples are given on the basis of €-scales. If the method can be conducted for non-continuous scales as well, this fact will be discussed at the end of the particular method's presentation. If the utility function is not monotonic, the determination has to be conducted separately for monotone subzones (as presented in Section 5.2.5).

9.4.2 Bisection version of the variable certainty equivalent method

The bisection version is an approach in which the decision maker has to determine certainty equivalents of lotteries. It is similar to the bisection method, which you have gotten to know in the determination of one-dimensional value functions.

In the first step the decision maker is presented with the basic-reference-lottery (x_{min}, 0.5; x_{max}, 0.5). This lottery's certainty equivalent is called $x_{0.5}$ and it holds that $u(x_{0.5}) = 0.5$ due to (9.19). In Figure 9-14 this first step is displayed in a). Analogous to the procedure of the bisection method, the intervals $[x_{min}, x_{0.5}]$ and $[x_{0.5}, x_{max}]$ are now "bisected" in terms of utility. For this purpose, one asks the decision maker for the certainty equivalents of the reference lotteries that are depicted in Figure 9-14 b) and c). These certainty equivalents are called $x_{0.25}$ and $x_{0.75}$ respectively, and it holds that $u(x_{0.25}) = 0.25$ and $u(x_{0.75}) = 0.75$.

A consistency check, again analogous to the procedure for the determination of value functions, must be attached to the determination of the nodes for a utility function. For this, the easiest possibility is to ask the decision maker for the certainty equivalent of the lottery ($x_{0.75}$, 0.5; $x_{0.25}$, 0.5). In case all questions were responded to consistently, the center in terms of utility of the interval $[x_{0.25}, x_{0.75}]$ must be the value $x_{0.5}$.

The fact that – in this method – all lotteries have simple 50-50 probabilities has to be looked upon favorably. As has been shown in previous chapters, decision makers have a hard time processing probabilities. The 50-50 lotteries appearing in this method, however, are the easiest lotteries and are made plausible to the decision maker by events such as a coin toss. Furthermore, the simple way of checking consistency has to be emphasized. If the survey shows, that further values need to be determined, then the intervals that were determined so far – or a subset of them – can again be bisected in terms of utility by simple 50-50 lotteries. The method's disadvantage is that any round of the survey contains the results of previous rounds. If a decision maker, for instance, made a mistake in the declaration of the value $x_{0.5}$ then this mistake is further propagated in the next questioning steps. In the case of a repeated bisection of the utility interval, systematic distortions can be amplified considerably.

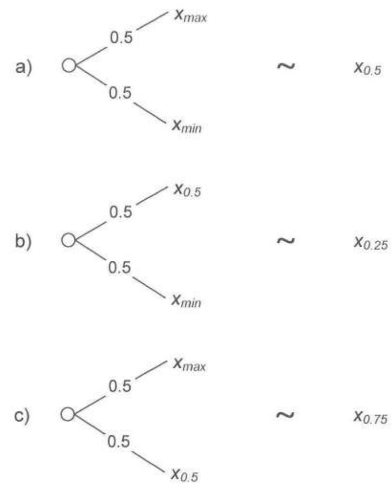


Figure 9-14: RL of the bisection version

We want to exemplify the method. A decision maker wants to determine his utility function for the interval [€0, €1,000]. In the first step, he is presented with the lottery (€0, 0.5; €1,000, 0.5). Assume that the decision maker is risk averse and names €400 as his certainty equivalent for this lottery. In the next step, the decision maker is presented with the lotteries (€0, 0.5; €400, 0.5) and (€400, 0.5; €1,000, 0.5). He assesses these lotteries with the certainty equivalents €180 and €600. In the consistency check, he is asked for the certainty equivalent of the lottery (€180, 0.5; €600, 0.5). If he does not name the amount of €400 as the certainty equivalent then this inconsistency must – again completely analogous to the determination of the value function – be eliminated. As explained in Chapter 5, the inconsistency can be eliminated by repeated questioning, it can be ignored by averaging or – as will be explained later in more detail – it can be represented directly by the concept of incomplete information. After a successful consistency check, the decision maker's statements also give reason for further inquiry. The certainty equivalent of €400 for the lottery (€180, 0.5; €600, 0.5) implies risk proclivity for this section. It is now to be questioned to what extent the decision maker is able to reconcile this with his generally risk averse attitude that was mentioned in the beginning.

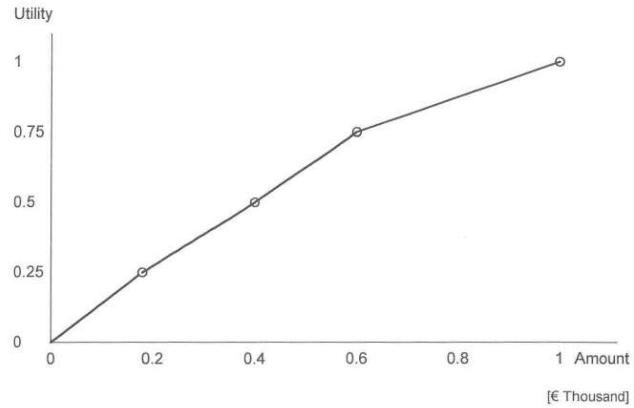


Figure 9-15: Utility function according to the bisection method

The five determined points of the utility function are drawn in a diagram and are connected to the utility function by piecewise linear interpolation or by adjustment of a given curve type. Figure 9-15 displays the result of a survey; it shows the utility function of the decision maker.

9.4.3 Quantile version of the variable certainty equivalent method

The quantile version of the variable certainty equivalent method only uses basic-reference-lotteries: the consequences x_{min} and x_{max} remain unchanged and the probabilities vary in each question. If, for example, four nodes of the utility function are to be obtained, the decision maker must name the certainty equivalents of the basic-reference-lottery for $p = 0.8$, 0.6, 0.4 and 0.2. For this method, the lotteries that have to be evaluated are shown in Figure 9-16.

The chosen probabilities in the basic-reference-lottery depend on the desired number of nodes. They do not have to necessarily be equidistant as in Figure 9-16. Here the trained decision maker can, depending on the particular decision problem, use his degrees of freedom wisely. From equation (9.19), the utilities of the certainty equivalents can be calculated in the quantile version more easily than in the bisection version; it holds that $u(x_{0.8}) = 0.8$ and so forth. For this method, a consistency check should be conducted as well – as for all measuring methods described in this book. A combination of the quantile method with other methods can be used for the consistency check. This way, the center of intervals in terms of utility could be determined, for instance the center of the interval $[x_{0.4}, x_{0.8}]$.

The quantile method certainly has the advantage that the consequences remain constant during the entire questioning method. Moreover, no statements from previous rounds are used as a foundation for further questioning steps. It is, however,

disadvantageous that not only 50-50 lotteries are considered. Even though only four probabilities appear in Figure 9-16, this method still poses relatively high demands on the information processing capabilities of the decision maker. Especially in the light of the distortions in probability judgments, as they were brought up in Chapter 7 and will be elaborately discussed in Chapter 13, one can by no means refrain from a consistency check for this method.

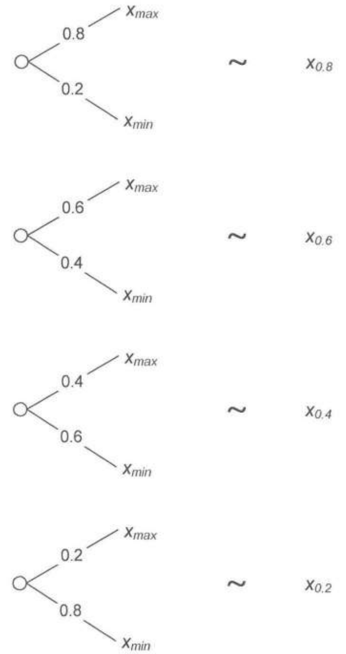


Figure 9-16: Basic-reference-lottery of the quantile version

9.4.4 Variable probability method

We want to introduce the variable probability method, a probability-equivalent-method, as the third method. In this method, the decision maker is presented with default consequences of the basic-reference-lottery as well as a default certainty equivalent and has to name the probability for which he is indifferent between the basic-reference-lottery and the certainty equivalent. The utility of the certainty equivalent equals the inquired probability. Preferably equidistant values between x_{min} and x_{max} are given for the certainty equivalents. If one wants to determine three nodes by questioning, the paired comparisons displayed in Figure 9-17 can be presented to the decision maker.

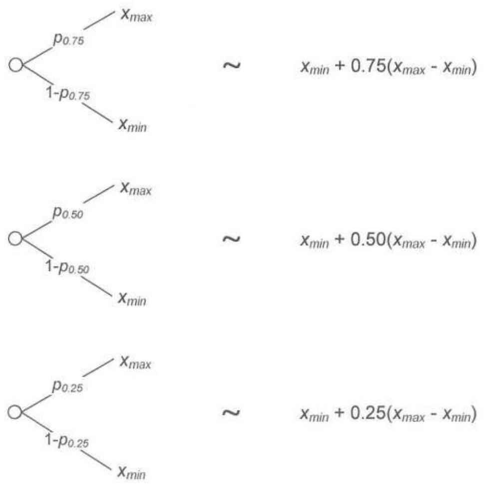


Figure 9-17: Basic-reference-lottery of the variable probability method

In this method, consistency checks should also be conducted. The variable probability method is often regarded as difficult. If decision makers are not sufficiently familiar with the concept, they will have difficulties naming indifference probabilities. On the contrary, the method has the advantage that no results of previous steps are taken into account for the new survey. In addition, a fairly special advantage of the method is that it can also be applied when the consequences are defined on non-continuous scales. In the two methods that were presented first, it could very well be the case that the scale on which the consequences are measured does not contain a certainty equivalent for the given lottery. Now examine, for instance, the dilemma of a coffee drinker, who can sweeten his coffee only with (whole) pieces of lump sugar and prefers more sugar to less sugar in the range of zero to three pieces. If you wish to obtain the utility function over the number of lump sugar pieces and you present the lottery (0 pieces, 0.5; 3 pieces, 0.5) to the decision maker (this lottery states, that you have a 50-50 chance to get a cup of coffee sweetened with zero or three pieces of sugar), then one of the authors

would like to name the number 1.3 as certainty equivalent, which is not possible according to the scale's definition.

9.4.5 Method of equal utility differences

Another method whose concept is already known from the determination of value functions in Chapter 5 (where it was called the method of equal value differences) is the method of equal utility differences. Here, the decision maker produces an increasing sequence of consequences by stating certainty equivalents; these consequences should all have the same difference in utility.

First, the decision maker chooses two consequences x_0 and x_1 such that $x_0 = x_{min}$ and the distance between x_0 and x_1 is roughly one-fifth of the length of the interval $[x_{min}, x_{max}]$. The decision maker is then asked to state a consequence x_2 such that he is indifferent between the lottery $(x_2, 0.5; x_0, 0.5)$ and the sure payment of x_1. As it then holds true that $u(x_1) = 0.5 \cdot u(x_2) + 0.5 \cdot u(x_0)$, we immediately conclude that $u(x_2) - u(x_1) = u(x_1) - u(x_0)$, i.e. the utility difference between x_2 and x_1 is the same as the utility difference between x_1 and x_0. Similarly, we ask for the consequence x_3 via the indifference statement $x_2 \sim (x_3, 0.5; x_1, 0.5)$. x_3 then lies one additional utility step of the same size above the utility of x_2. The analogy to the method of equal value differences in Chapter 5 should now be obvious. Similarly, for the method of equal utility differences we cannot expect that the produced sequence of consequences exactly hits the maximum outcome x_{max}. However, we already know how to solve this problem; the final normalization of the utility function simply occurs over a slightly larger interval containing x_{max}.

9.4.6 Trade-off method for utility functions

All previously described methods for determining utility functions suffer from the same issue; they assume that the decision maker's answers to questions about certainty equivalents to simple lotteries etc. can be interpreted in a sense that is in line with expected utility. In Chapter 13, we will see that this assumption is critical and that decision makers usually give distorted answers even if only simple lotteries are involved. At this point, however, we would like to stress a problem that is important to all utility elicitation methods discussed so far. When making intuitive decisions (like when answering questions for certainty equivalents to simple lotteries), small probabilities have a larger impact on the evaluation of a lottery than expected utility theory prescribes. Other probabilities are systematically over- or under-weighted as well. Neglecting these distortions results in systematically distorted utility functions. Bleichrodt et al. (2001) analyze how one can adjust answers that suffer from systematically distorted probabilities and hence elicit undistorted utility functions. The exact approach, however, is beyond the scope of this book and we refer the reader to the original paper.

An alternative is to apply methods which do not suffer from these biases. The trade-off method (which has nothing in common with the trade-off method from Chapter 6 except for the name) by Wakker and Deneffe (1996) is one of these methods. Compared with previous methods, it is more complicated. This is however more than offset as distorted probabilities no longer result in a distorted utility function.

The basic idea of the trade-off method for eliciting utility functions is similar to the method of equal utility differences. The decision maker produces a sequence of consequences which all have the same utility difference. While the method of equal utility differences breaks down if the probability $p = 0.5$ is systematically distorted (e.g. is treated like a probability of 0.4), the trade-off method is immune to such a bias.

The trade-off method requires two consequences x_a and x_b with $x_a \succ x_b$ to be chosen such that they – in the best case – lie outside of the interval $[x_{min}, x_{max}]$ of relevant consequences (the two consequences x_a and x_b are only needed for comparison reasons; we assume they are less than x_{min}).

Again, we set $x_0 = x_{min}$ and the decision maker is asked for a consequence x_1 that makes him indifferent between the lotteries $(x_0, p; x_a, 1-p)$ and $(x_1, p; x_b, 1-p)$. Presumably, a probability of $p = 0.5$ will make this question particularly easy to answer. However, the trade-off method works with every other p. Some algebra shows that this indifference statement implies $u(x_1) - u(x_0) = p/(1-p)\cdot(u(x_a) - u(x_b))$. The utility difference between x_1 and x_0 is given by the value $p/(1-p)\cdot(u(x_a) - u(x_b))$. This value is unknown and furthermore it is affected by the potentially distorted probabilities p and $(1-p)$. However, it is not necessary to know this value of the utility difference because the next step delivers a similar insight. We ask the decision maker for a consequence x_2 that makes him indifferent between $(x_1, p; x_a, 1-p)$ and $(x_2, p; x_b, 1-p)$. From this indifference, we deduce that the utility difference between x_2 and x_1 is also given by the same value $u(x_2) - u(x_1) = p/(1-p)\cdot(u(x_a) - u(x_b))$. In particular we conclude that $u(x_2) - u(x_1) = u(x_1) - u(x_0)$ and this holds true independent of the exact choice of the potentially distorted probability p. The next steps are obvious now: we produce a whole series of indifference statements $(x_i, p; x_a, 1-p) \sim (x_{i+1}, p; x_b, 1-p)$ which all result in equal utility indifferences. A convenient choice of x_a and x_b (the closer they are, the smaller the distances in the sequence of consequences become) makes it possible for us to reach (or exceed) the consequence x_{max} within four to five steps, as in the method of equal utility differences. The normalization of the utility function is then as usual.

We want to illustrate the trade-off method with an example. Assume a decision maker wants to know his utility function on the interval $[x_{min} = €1{,}000, x_{max} = €10{,}000]$. We set $x_a = €100$ and $x_b = €500$, so that these two consequences lie outside of the previously specified interval. The next steps are shown in Figure 9-18. In a first step, we ask for a consequence x_1 that makes the decision maker indifferent between the lotteries (€1,000, 30%; €100, 70%) and (x_1, 30%; €500, 70%). As already noted, we expect a probability of $p = 0.5$ to ease the whole procedure for the decision maker. For didactical reasons, however, we here chose $p = 0.3$. Assume the decision maker states $x_1 = €2{,}500$. Then, in a second step, we ask for x_2 that makes him indifferent between the lotteries (€2,500, 30%; €100, 70%) and (x_2, 30%; €500, 70%). We obtain a further value, say, $x_2 = €6{,}000$ that we can use for the next step. Assume the decision maker reveals $x_3 = €11{,}000$; we then stop the elicitation process as €11,000 exceeds $x_{max} = €10{,}000$.

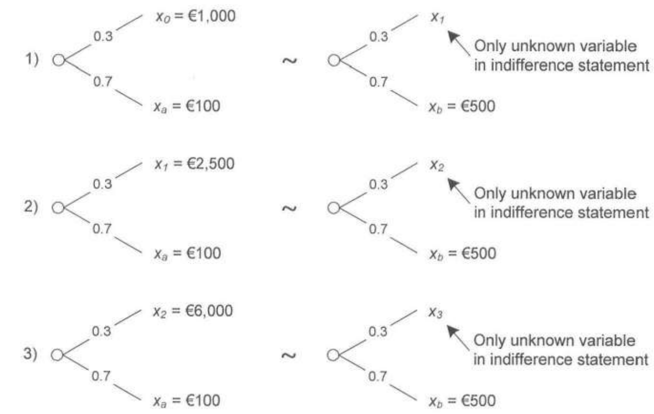


Figure 9-18: Lotteries and indifference statements in the trade-off method

The resulting points on the decision maker's utility function are shown in Figure 9-19. To facilitate the understanding, we have plotted two utility axes: on the left hand side, a non-normalized, arbitrarily scaled utility axis and on the right hand side a utility axis normalized on the interval $[x_{min} = €1{,}000, x_3 = €11{,}000]$. You can see that, on the non-normalized utility axis, the scaled utility difference $(1-p)/p\cdot[u(x_b) - u(x_a)]$ resulting from the utility difference $u(x_b) - u(x_a)$ and our choice of the lottery's probability $p = 0.3$ determine the equal utility difference between all three elicited utility values $u(x_1)$, $u(x_2)$ and $u(x_3)$. The well-known normalization then leads to the utility values $u(x_1) = 1/3$, $u(x_2) = 2/3$ and $u(x_3) = 1$ as is depicted on the normalized utility axis on the right hand side.

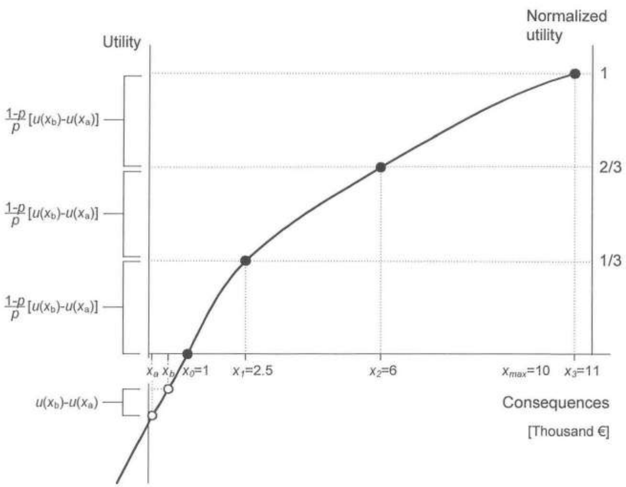


Figure 9-19: Utility function elicited with the trade-off method

9.4.7 Consistency check

Now that we have presented the methods, you should practice them with practical examples. In doing so, you will notice that the utility function can be different for the same decision problem and the same decision maker depending on the method of determination. Considering that the determination of utility functions is about the measurement of preferences and that measuring methods can cause errors and systematic distortions, you will not be surprised by the different utility functions. Inconsistencies can not only occur within a method, but also between methods. In certain methods, especially in those that use small probabilities and do not make attempts to avoid distortions like the trade-off method does, strong systematic distortions can even occur (Hershey et al. 1982, Hershey and Schoemaker 1985). To avoid these distortions, the utility function should be determined with different methods. Analogous to the approach for the determination of a utility function with only one elicitation method, inconsistencies between two or more methods should be pointed out to the decision maker. Inconsistencies should be compensated by averaging or one should continue by use of the concept of incomplete information (cf. Section 10.2.2). This feedback and the questioning with different methods can be carried out especially well by use of an interactive computer program (von Nitzsch and Weber 1986).

A thorough determination of the utility function is an art. Decision makers normally do not have their utility function readily accessible in their minds. They

often establish their preference during the interview; the way of questioning can hence easily have some influence on the determined utility function.

For the prescriptive application of utility theory, it is important that the problems related to behavioral science should not lead to a rejection of the theory; they rather show that decision makers cannot unaidedly fulfill rationality claims in simple – and even less in complex – decision situations. The descriptive findings on systematic distortions are quite significant for prescriptive decision science since they demonstrate an intensified need for decision support and point out the necessity of methods to avoid such distortions. The development of the trade-off method for eliciting utility functions is a good example here. Without descriptive findings about the evaluation and perception of (distorted) probabilities, this – admittedly rather complicated – method would not have been invented. The trade-off method's complexity, however, is more than offset by its immunity against distorted probabilities.

9.4.8 Determination of the utility function on the basis of the decision maker's risk attitude

In addition to direct questioning with the aid of reference-lotteries, the decision maker's risk attitude can also be used to determine the utility function. As derived before, a risk-averse decision maker must have a concave utility function and a risk-prone decision maker must have a convex utility function. If the decision maker knows that he has a monotonically increasing utility function and that he is risk-averse, the admissible range for the utility function can be strongly limited by only one single question. Figure 9-20 clarifies the following argumentation.

If the point $(x, u(x))$ was determined by questioning the decision maker, a monotonically increasing, concave utility function can only be located in the shaded area. By clever choice of a few consequences, the admissible range of the utility function can in addition be strongly limited due to the decision maker's general statement of risk aversion.

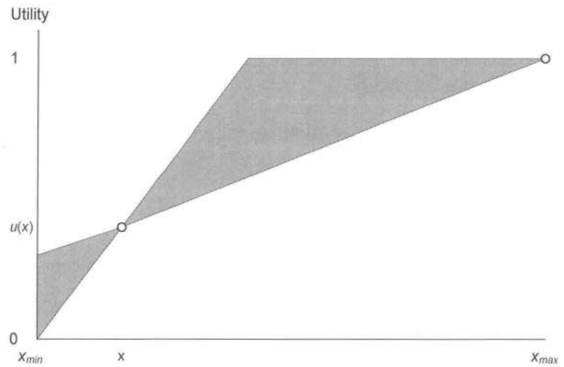


Figure 9-20: Possible range of the utility function

The utility function can possibly be determined more easily if more precise statements about the risk attitude are available. In sections 9.3.3 and 9.3.4, we have gotten to know the concepts of absolute and relative risk aversion and simultaneously saw that certain classes of utility functions are characterized by a constant (absolute and relative) risk attitude. If one knows that a decision maker has constant risk attitude, only the free parameters of the corresponding class of utility functions have to be determined. We want to explain this approach for the case of constant absolute and constant relative risk aversion.

Now let us first assume that the decision maker has constant absolute risk aversion (CARA) – i.e., that his Arrow/Pratt measure $r(x)$ is constant. Whether this is indeed the case can for example be checked by use of the certainty equivalent criterion that we discussed in Section 9.3.4. If one adds a quantity δ to each consequence, the new lottery's certainty equivalent has to equal CE + δ. The constancy of the risk measure can quite easily be checked for some chosen lotteries by variation of δ and by direct determination of the certainty equivalents. As described earlier in Section 9.3.4, for CARA the decision maker's utility function is:

$$u(x) = \alpha + \beta e^{-cx} \quad \text{with} \; c = r(x) > 0 \; \text{and} \; \beta < 0 . \tag{9.21}$$

Since α and β are scaling constants which scale the utility function on the interval [0, 1], only the parameter $c = r(x)$, the constant Arrow/Pratt risk measure, has to be determined. To accomplish this, a basic-reference-lottery can be presented to the decision maker for which he has to determine the certainty equivalent or, if that was given as well, a quantity of the basic-reference-lottery so that he is indifferent between certainty equivalent and basic-reference-lottery. From this indifference statement the parameter c (= $r(x)$) can be derived.

We want to present two possibilities for the determination of the parameter c. The first one requires the indication of a certainty equivalent for a 50-50 lottery, the second and mathematically easier method requires the somewhat harder indication of an indifference probability.

With the first method, we assume that a decision maker has a constant Arrow/Pratt measure and he stated his certainty equivalent to be €3.80 for the lottery (€10, 0.5; €0, 0.5). It holds:

$$\begin{gathered} \text{EU(lottery)} = u(CE), \text{ i.e} \\ 0.5(\alpha + \beta e^{-c\cdot 10}) + 0.5(\alpha + \beta e^{-c\cdot 0}) = \alpha + \beta e^{-c\cdot 3.8}. \end{gathered} \tag{9.22}$$

(9.22) is equivalent to $0.5\ e^{-c\cdot 10} + 0.5 = e^{-c\cdot 3.8}$. An approximate solution of this equation using methods included in spreadsheet programs results in $c = 0.1$. One can compute the coefficients α and β from the following two equations:

$$\begin{gathered} \alpha + \beta\, e^{-c\cdot 10} = 1 \\ \alpha + \beta\, e^{-c\cdot 0} = 0. \end{gathered} \tag{9.23}$$

We get $\alpha = 1.58$ and $\beta = -1.58$. In order to simplify the computation in the determination of parameter c, one can also prepare a chart that indicates the value of c depending on the risk premium. It holds for the lottery above:

$$\begin{gathered} \text{RP} = 0.62 \Rightarrow c = r(x) = 0.05 \\ \text{RP} = 1.2 \Rightarrow c = r(x) = 0.1 \\ \text{RP} = 2.15 \Rightarrow c = r(x) = 0.2 \\ \text{RP} = 3.63 \Rightarrow c = r(x) = 0.5 \\ \text{RP} = 4.31 \Rightarrow c = r(x) = 1. \end{gathered}$$

For the second method for the determination of c, one assumes directly the standardized form of the exponential utility function instead of assuming (9.21):

$$u(x) = \frac{1 - e^{-c\frac{x - x_{min}}{x_{max} - x_{min}}}}{1 - e^{-c}}. \tag{9.24}$$

The decision maker is then presented with the BRL (x_{min}, p; x_{max}, $1 - p$) and he is asked to state the probability p for which he is indifferent between the BRL and the safe consequence $x = 0.5\ (x_{min} + x_{max})$. The desired constant c of the utility function can be deduced directly from the indifference probability. It holds that:

$$c = 2 \cdot \ln(\frac{1}{p} - 1). \tag{9.25}$$

If not the absolute risk attitude but instead the relative risk attitude $r^*(x) = x \cdot r(x)$ is constant, the utility function again can be determined directly in an analogous way. See section 9.3.4 for a description of what CRRA utility function looks like (equation (9.16)).

Even though the function looks somewhat more complicated, only the constant quantity $x \cdot r(x)$ has to be determined, as in the case of constant absolute risk aversion (CARA), by an indifference statement. An outline of further connections between risk attitude and the shape of the utility function can be found in Harvey (1981).

9.5 Computation of the optimal alternative

After the utility function has been determined, the alternatives can be put in order for a given decision problem. This way, the optimal alternative can of course be determined as well. The kind of determination depends on the chosen display format of the alternatives.

If the alternatives are given in the form of a decision matrix or in the lottery format, the expected utility of the alternative can easily be computed. The expected utility equals the sum of the products of the consequences' utilities and the probabilities for the consequences' occurrence. As a reminder, the formula 9.3 is repeated:

$$\mathrm{EU}(a) = \sum_{i=1}^{n} p_i \cdot u(a_i) \tag{9.26}$$

This display format implies a single-stage decision situation and a finite number of consequences. An example from Chapter 2 shall serve to demonstrate this case. The decision matrix in Table 9-3 contains profits in thousands of euros.

Table 9-3: Decision matrix with consequences evaluated in thousands of euros

s_i	s_1	s_2	s_3	s_4	s_5	s_6
$p(s_i)$	0.10	0.15	0.15	0.30	0.20	0.10
a	0	15	15	15	15	15
b	–20	–5	10	25	25	25
c	–40	–25	–10	5	20	35

After the consequences have been evaluated with the utility function (CARA with $c = 0.02$):

$$u(x) = 1.287 - 0.578 \cdot e^{-0.02x} \tag{9.27}$$

the decision matrix reads as in Table 9-4.

Table 9-4: Decision matrix with consequences evaluated in utility

s_i	s_1	s_2	s_3	s_4	s_5	s_6
$p(s_i)$	0.10	0.15	0.15	0.30	0.20	0.10
a	0.71	0.86	0.86	0.86	0.86	0.86
b	0.43	0.65	0.81	0.94	0.94	0.94
c	0.00	0.33	0.58	0.76	0.90	1.00

The expected utilities of the alternatives amount to:

$$\mathrm{EU}(a) = 0.845$$
$$\mathrm{EU}(b) = 0.825$$
$$\mathrm{EU}(c) = 0.645.$$

A rational decision maker would hence choose alternative a.

Infinite consequence sets – as for instance defined by the normal distribution - are prevalent in many applications. Capital market theory generally assumes that stock returns are normally distributed. In case of a continuous distribution, the expected utility of a lottery can be written as:

$$EU(a) = \int_{-\infty}^{+\infty} u(x) f(x) dx \tag{9.28}$$

where $u(x)$ represents the utility function and $f(x)$ represents the density function of the distribution of the consequences of alternative a. For continuous distributions and piecewise linear utility functions, the expected utility can usually be calculated easily. Often the computation is also possible if the functional shape of the utility function as well as the density function is given. In the literature, tables can be found that indicate the expected utility for certain types of probability distributions and certain shapes of the utility function. Relatively easy expected utility formulas arise for the case of the exponential utility function (utility function with constant absolute risk aversion, cp. Keeney and Raiffa 1976, p. 202). The certainty equivalent (and hence also the expected) can be determined especially easy in case the consequences are normally distributed and the decision maker has an exponential utility function with risk attitude parameter c. This is also referred to as the so-called "hybrid-approach" (Bamberg 1986). In this case it holds that:

$$\mathrm{CE}(a) = u^{-1}(\mathrm{EU}(a)) = \mathrm{EV}(a) - 1/2 \cdot c \cdot \mathrm{Var}(a). \tag{9.29}$$

In a multi-stage model, as can be represented by a decision tree, the objective of the considerations must be to find the optimal multi-stage alternative (=strategy). For cases of multi-stage decision problems, we have distinguished two display formats. If the strategies are given in the form of a decision matrix, the determination of the optimal strategy is equivalent to the determination of the optimal alternative in the single-stage model. As is made clear in the example above, one hence calculates the expected utility of the strategies, orders the strategies and chooses the optimal strategy. However, we have seen before that it can be desirable to represent the multi-stage decision problem in the form of a decision tree. The procedure is then as follows:

In the decision tree, the strategy with maximal expected utility can be determined by the *roll-back-method*. The procedure is as follows:

1. The consequences are first evaluated using the utility function.
2. Starting from the consequences, one moves to the preceding decision node.
3. Here, the expected utility of all alternatives given at this decision node is calculated. The alternative with the highest expected utility is determined and all others are eliminated.
4. After all decision nodes of the last stage were dealt with in this way, one proceeds the same way for the penultimate stage. Arriving at the first decision node, the optimal alternative with the highest expected utility is then determined.

The *roll-back-method* shall be illustrated with an example. For that purpose, we resort to the oil drill-problem which is prevalent in literature. It is to be decided whether a seismic test should be made before an oil drill, whether drilling should be conducted directly or whether it should be decided not to drill at all. Figure 9-21 displays the decision tree for this problem with the consequences indicated in thousands of euros as well as in utility values, which result from an assumed utility function. This utility function is defined over the interval [–€130,000, €270,000].

In the *roll-back-method*, the consequences have to be examined backwards, until a decision node is reached. We begin with the decision node obtained if a test was conducted and the test result was beneficial. If drilling will not occur, a utility of 0.641 is certain; this corresponds to a loss of €30,000 because of the test costs. The alternative "drill" has an expected utility of $0.85 \cdot 0.995 + 0.15 \cdot 0 = 0.846$; the alternative "drill" is chosen as it yields the higher expected utility and the alternative "no drill" is eliminated, as displayed in Figure 9-21. In this fashion, the entire decision tree is completed backwards. The numbers in the boxes signify expected utility values. It can be seen that the strategy "seismic test, drill in case the result is beneficial, do not drill in case the result is disadvantageous" has the highest expected utility with 0.764, followed by the strategy "do not drill" with 0.738.

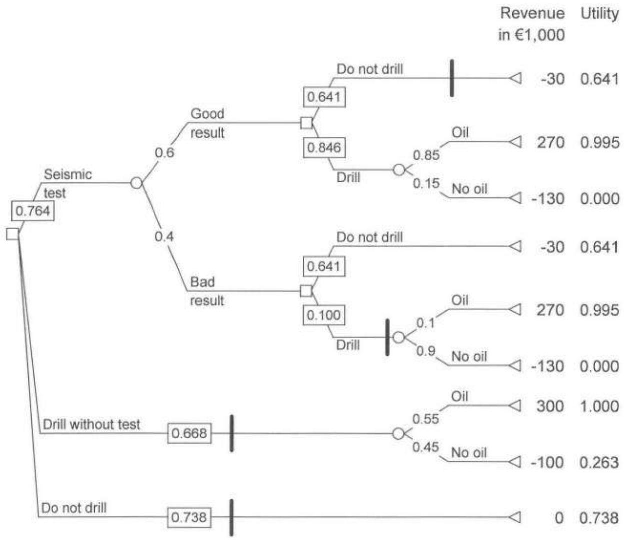


Figure 9-21: Decision tree for the oil drill-problem

The *roll-back-method* cannot be conducted for every arbitrary preference theory. However, it can be shown for expected utility theory, that the elimination of alternatives with lesser expected utility does not falsify the global optimum. This property in utility theory is essentially based upon the validity of the independence axiom. If probabilities could not be multiplied (*reduction of compound lotteries axiom*) or if irrelevant branches of the decision tree could not be neglected (independence axiom), the *roll-back-method* would be invalid.[2]

In the *roll-back-method*, only the optimal strategy in terms of the maximal expected utility is determined. The decision maker does not receive any information about all other possible strategies. If one wants to compare a second best strategy to the best strategy in the way of a sensitivity analysis (which will be discussed in the next chapter) the *roll-back-method* cannot serve as an adequate selection method. All strategies that need to be considered in this case would have to be presented within a decision matrix and would have to be evaluated with the expected utility criterion.

[2] One example for a problematic decision calculus in this respect is the cost-benefit ratio (cf. Case study E in Eisenführ et al. 2001).

9.6 Utility theory and risk

9.6.1 Connection between value and utility function

Can we now truly characterize the risk behavior of decision makers with the utility function, for instance with the aid of the Arrow/Pratt measure? Unfortunately, the term "risk averse" is not unambiguous, i.e. it does not necessarily imply that a decision maker in all situations "avoids risk". The reason for the concavity of the utility function can lie in the fact that the decision maker obtains only decreasing marginal value (in terms of a value function) from increasing safe values of the objective variable; this would suffice for his value function of this objective variable to be concave. The utility function's concavity can also be due to the decision maker's dread of risky situations. A precise statement about the risk attitude would only be obtained if the risk attitude were measured relative to the value function. Assume that the following relation between the value function and the utility function of a decision maker holds: $u(x) = \mathrm{f}(v(x))$. Statements about the relative risk behavior of this decision maker could be derived from a comparison of the curvature of the utility function and the value function (where "relative risk behavior" means relative to the value function, not relative to final wealth like for the Arrow/Pratt measure). Whenever we look at the risk behavior of a decision maker relative to the value function in the following, we want to talk about his *intrinsic* risk attitude. An example shall serve for explanatory purposes (based on Dyer and Sarin 1982).

Assume that you are indifferent between two oranges and the lottery (0 oranges, 0.5; 8 oranges, 0.5), and your utility function on the interval [0 oranges, 8 oranges] is monotonically increasing. You would thus be classified as risk averse since your certainty equivalent is smaller than the expected value of the lottery.

Let us now initially assume that you are indifferent between the safe transitions (0 oranges → 2 oranges) and (2 oranges → 8 oranges). In this case, the risk premium can be explained by the value function: measurable value function and utility function are identical for the considered points. You are risk neutral relative to the value function, i.e. you are intrinsically risk neutral.

Let us alternatively assume that you have a linear value function on the interval [0 oranges, 8 oranges], i.e. you are indifferent between (0 oranges → 4 oranges) and (4 oranges → 8 oranges). In this case, the risk aversion cannot be explained by the value function. The utility function is "more concave" than the value function, i.e. you are intrinsically risk averse. This case is once again clarified in Figure 9-22.

We want to only briefly refer to papers that theoretically investigate the intrinsic risk attitude of decision makers (Krelle and Coenen 1968, Dyer and Sarin 1982, Sarin 1982, Wilhelm 1986, Kürsten 1992a and 1992b). Let us mention the work of Smidts (1997) who empirically determines the utility and value functions of 200 Dutch farmers. He is able to show that these functions differ significantly

from each other and that the relation between both functions can best be described by an exponential function.

The fact that risk and value conception in utility calculus are inseparably connected has evoked great uncertainty, especially in the German literature. At this point, we can see no necessity to pick up the discussion. The questions that have been raised are answered clearly in the papers that were previously quoted. A nice, concluding examination can be found in Bamberg et al. (2008) and Dyckhoff (1993).

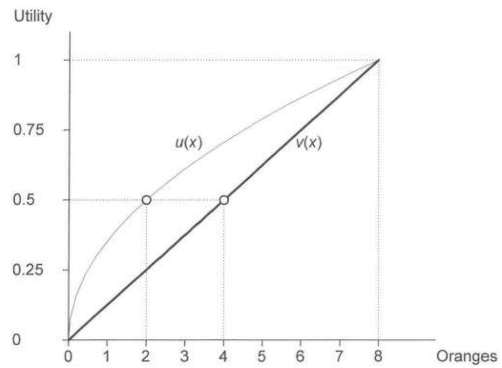


Figure 9-22: Utility and value function for intrinsic risk aversion

9.6.2 Risk definition for equal expected value of lotteries

So far we have approached the objective of learning something about the risk of a lottery by just a small step. Rothschild and Stiglitz (1970) simplify the problem by only considering lotteries with identical expected values.

They give three alternative definitions for the relation "lottery a is riskier than lottery b". They prove the equivalence of the definitions, meaning that they can all be applied in place of each other. The definitions are: a lottery a is riskier than a lottery b if and only if:

1. every risk averse decision maker (in terms of utility theory) prefers a over b;
2. a can be derived from b by adding a random variable with an expected value of zero to each possible outcome of b;
3. a was derived from b by a *mean preserving spread*. In a *mean preserving spread*, elements are taken from the center of the distribution of b and are transformed to the edge of the distribution without this transformation changing the expected value of b.

The relation gained by the three equivalent definitions "lottery a is riskier than lottery b" is strongly based on utility theory. The relation is not equivalent to the statement "a has a greater variance than b"; variance is no measure for a lottery's risk in terms of utility theory. Alternatively, expressed somewhat more formally, the risk premium only approximately equals half of the variance multiplied by the Arrow/Pratt measure.

Consider for instance the three lotteries:

$$a = (€30, 0{,}5; €10, 0{,}5)$$

$$b = (€25, 0{,}8; €0, 0{,}2)$$

$$c = (€40, 0{,}2; €15, 0{,}8).$$

All lotteries have identical expected value (€20) as well as identical variance ($€^2 100$). Nevertheless, a risk averse decision maker whose preference is given for instance by the utility function $u(x) = 1 - e^{-0.1x}$, should have the following preference order: $c \succ a \succ b$.

9.6.3 Utility – A function of value and risk?

We want to continue trying to answer the question what the risk of a lottery is. The background of this question also contains the search for the possibility of understanding the expected utility of a lottery as a function of the expected value and risk of the lottery. In everyday language, alternatives are often evaluated by comparing the risk as the "bad" and the expected value as the "good" of the alternatives. For example, the expected profit of an investment is considered too small compared to the risk that needs to be borne. The theoretically correct expected utility is (very) rarely used in the line of reasoning. It is hence important to understand the connection between calculi that are based on expected value and risk on the one side and those based on utility theory on the other side. Two approaches to analyze the connection can be distinguished (cp. in more detail on the approaches and on this section Sarin and Weber 1993a).

One possibility consists in separately determining risk and value of lotteries and then composing preference judgments of both components. The decision maker is asked *directly* about risk assessment and value estimation. In doing so, a definition of risk and value is not needed – merely the decision maker's perception regarding risk and value is of interest. If the order of the lotteries regarding risk and the order regarding value fulfill some general properties, this order can be represented by a risk function and a value function. This approach requires the decision maker to actually order lotteries concerning risk and concerning value. Empirical studies have shown that this precondition is fulfilled (cp. for example Keller et al. 1986 or the overview on risk perception in Weber 1990). More general risk statements about the risk of nuclear power, flying and so forth can be measured reliably as well (cp. Slovic 1987). Burgemeister and Weber (1993) have analyzed how respondents perceive the risk of new technologies and determined that decision makers are able to make risk assessments here as well. You can find an

overview about approaches on risk measurement in Brachinger and Weber (1997). The derivation of the preference discussed here has the disadvantage that the preference, composed of risk judgments and value judgments, does not necessarily fulfill the rationality postulates of utility theory. Although interesting insight can be gained for descriptive purposes, this approach cannot (yet) convince from a prescriptive point of view based on expected utility theory.

The *second* possibility consists of generating a preference functional equivalent to the utility of a lottery linking the expected value to a second quantity which is usually defined as the lottery's risk. The preference, represented by the lottery's utility, is the output variable here. The preference is decomposed into two components (expected value and risk).

In what follows, we concentrate on the case especially important for economists in which the expected utility of a lottery a is based on the expected value and the variance (Jia and Dyer 1996 provide a more general approach).

To guarantee the rationality of the preference functional, it must hold that:

$$\mathrm{EU}(a) = f(\mathrm{EW}(a), \mathrm{Var}(a)). \tag{9.30}$$

Risky alternatives are represented by expected value and variance in many areas of business economics. Modern capital market theory is based on this, marketing strategies are characterized in this way, and this approach has also attracted attention in strategic planning. The expected value-variance-rule is also called the (μ,σ)-rule because of the two dimensions, namely expected value and variance or expected value and standard deviation. An alternative can be represented in a two dimensional diagram.

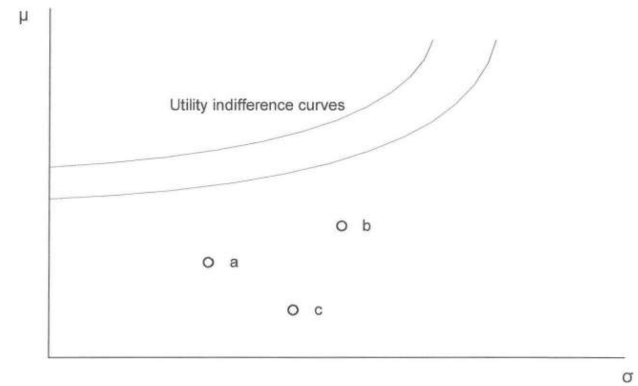


Figure 9-23: (μ,σ)-diagram

Three alternatives *a, b, c* are depicted in Figure 9-23. Alternative *c* is dominated for risk averse decision makers because it has a higher risk and a smaller expected value compared to *a*. This is referred to as (μ,σ)-dominance. If a decision maker is risk averse in terms of Arrow-Pratt, he has a concave utility function. The utility *indifference* curves in the (μ,σ)-diagram are, however, convex for risk averse decision makers as displayed in Figure 9-23. A higher risk, measured by a higher variance or standard deviation, requires a higher expected value to yield the same utility. A risk neutral or slightly risk averse decision maker will choose alternative *b* in Figure 9-23, a more risk averse decision maker will prefer *a* over *b*.

In the following, we want to discuss the case in which a decomposition of a lottery's expected utility in expected value and variance is possible. Based on the considerations of Jia and Dyer (1996), we already know that a decomposition is not possible for arbitrary risky alternatives and utility functions. A certain class of utility functions or a certain distribution type of the consequences has to be available (Schneeweiß 1967).

For the special case of quadratic utility functions, the expected utility of a lottery can be written as a function of expected value and variance of the lottery. It holds that:

$$\mathrm{EU}(a) = \mathrm{EV}(a) - \alpha[\mathrm{EV}(a)^2 + \mathrm{Var}(a)], \quad \alpha > 0. \tag{9.31}$$

The corresponding quadratic utility functions read as follows:

$$u(x) = x - \alpha x^2. \tag{9.32}$$

The functions are monotonically increasing in the interval $[x_{min}, 1/2\alpha]$.

In order to establish the equivalence between expected utility theory and expected value-variance-approach, the amount of acceptable lotteries (i.e. the amount of acceptable probability distributions of consequences) can also be restricted. In case the consequences are normally distributed, it can be shown that the expected utility of the lottery is equivalent to a preference function based on expected value and variance.

Utility function and distribution can certainly be restricted simultaneously as well; this was outlined in Chapter 9.5 in the form of the hybrid approach.

The necessary conditions for the decomposition of expected utility into expected value and variance are fulfilled only in few cases. Therefore, you should take care whenever confronted with expected value-variance-rules in the course of your studies. As explained in Section 9.3.4, the quadratic utility function has the undesirable property that absolute as well as relative risk attitude increase with growing wealth. Furthermore, the discrete set of states considered in this book and the set of states prevailing in many real decision situations do not permit any normal distributions.

Questions and exercises

9.1

(a) What is the relationship between a value function and a utility function?

(b) Describe two methods to determine utility function and, if applicable, describe the analogous methods of determining value functions.

9.2

Annette, Lucas, and Martin each have a ticket for a tombola which pays out €100 with a probability of 0.2, and €10 with a probability of 0.8. The utility functions are defined as follows:

- Annette: $u(x) = 0{,}002x^2 + x$
- Lucas: $u(x) = \log x$
- Martin: $u(x) = 0{,}4x + 100$.

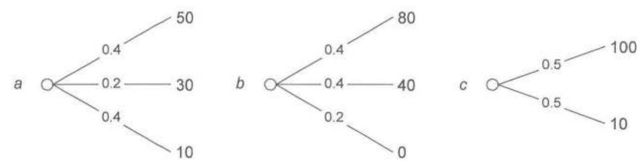
(a) Draw the three utility functions using the usual normalization to [1; 0].

(b) Someone wants to buy a ticket from them for €25. Who would participate in that deal?

(c) Define the risk premium for every person.

9.3

Lothar Lotter has to order these three lotteries according to his preference:



He asks the decision theoretician Bernd Nulli for help. The expert suggests ordering the lotteries according to the utility value. Lothar determines his utility function, which is $u(x) = x/50 - x^2/10{,}000$. To make sure he makes the right decision, Lothar enlists the help of the investment analyst Miller-Sigman, a proponent of the μ-σ-principle. He advises Lothar to decide depending on the expected value and the variance of each lottery. Miller-Sigman also assists Lothar in defining a μ-σ-preference function. It is $f(\mu,\sigma) = \mu/50 - (\mu^2 + \sigma^2)/10{,}000$.

Which preference orders are the results of the two approaches for Lothar? Please comment on the results.

9.4

Jack accidently damages Ute's vase which is of high but hard to estimate value. In order to settle the claim, Klaus's insurance makes Ute an offer of €10,000. Ute considers whether to accept the offer or to ask a lawyer to file a claim for €50,000. Ute assumes that the insurance will react by offering €25,000 or repeating the offer of €10,000; she thinks both alternatives are equally likely. Ute would accept

€25,000 if it were offered. If the offer of €10,000 is repeated she would have the choice of accepting it or taking legal action, again with an uncertain result. She thinks that the outcomes of €10,000, €25,000 and €50,000 are equally likely.

1. Depict the decision problem in a decision tree.
2. Show three strategies from which Ute can choose.
3. Which strategy would Ute pick if she had the following utility function:

$$u(x) = \sqrt{\frac{x}{50{,}000}}.$$

9.5
A decision-maker has constant relative risk aversion. His certainty equivalent for the lottery (€100, 0.5; €20, 0.5) is equal to €50. State a utility function which is in line with this for the interval [€20; €100] and normalize the values to lie between 0 and 1.

9.6
The number of job losses as a result of the upcoming reorganization of Lien PLC is still unsure but it is estimated that it could lie between 0 and 2,000. The works committee is indifferent between a 50-50 lottery of losing 2,000 or 0 jobs and a sure loss of 500 jobs. Other valid statements are $(0, 0.5; 500, 0.5) \sim 200$ and $(500, 0.5; 2{,}000, 0.5) \sim 900$.

(a) Draw a sketch and interpolate by fitting a curve.
(b) Does the curve show risk aversion or risk proclivity?
(c) You have to choose from three alternatives. For procedure a, 300, 600, and 1,000 job losses are estimated with a probability of 1/3 each. Procedure b means a loss of 500 jobs. For procedure c, 0 job losses are estimated with a probability of 0.25, 300 job losses with a probability of 0.5, and 2,000 job losses with a probability of 0.25. Which alternative does the works committee prefer to be consistent with its utility function?

9.7
Slovic and Tversky (1974) present arguments for and against the validity of the independence axiom to decision-makers. Do you know any?

9.8
For the next week, look out for the contexts in which the word "risk" appears. How can these statements be classified according to the discussion in section 9.3?

9.9
A decision-maker has the utility function $u(x) = \sqrt{x+a}$ and the value function $v(x) = \ln\sqrt{x+a}$ with $x > 0$, $a > 0$.

(a) Which is the absolute risk attitude of the decision-maker and how does it develop with increasing wealth x?

(b) Which is the risk attitude relative to his wealth x and how does it develop if the wealth increases?
(c) Which risk attitudes does the decision-maker have relative to his value function?
(d) Comment on the different results.

9.10
A decision-maker has a constant relative risk aversion of 0.5. Determine a matching utility function and normalize it to values between 0 and 1 for $x \in [0; 100]$.

9.11
Which method to determine the utility function was used in the following interview? Which advantages and disadvantages result from the different methods? Draw the utility function of the decision-maker and characterize their risk attitude.

(a) The interview of Alfred about his certainty equivalents for lotteries with two equally likely outcomes ($p = 0.5$) resulted in:

	1st lottery	2nd lottery	3rd lottery	4th lottery
x_{max}	2,000	800	350	2,000
x_{min}	0	0	0	800
Certainty equivalent	800	350	100	1,200

The following utility values are given:

$$u(2{,}000) = 1$$
$$u(0) = 0.$$

(b) Decision-maker Boris is asked for probabilities which make him indifferent between given basic reference lotteries and certainty equivalents:

$$x_{max} = 2{,}000 \qquad u(2{,}000) = 1$$
$$x_{min} = 0 \qquad u(0) = 0.$$

Certainty equivalent	400	800	1,200	1,600
$p^*(x_{max})$	0.15	0.3	0.5	0.7

9.12
The scrap rate in a tool production facility has increased dramatically in the last month. Currently, the fraction of scrap amounts to 20% and costs of €500,000. The high amount of scrap indicates an error in the production technology which can be corrected with 90% probability by stopping the production process. If the error is not corrected, it has to be assumed that the costs of the last period will continue to occur. With a probability of 10%, the high amount of scrap occurred only by chance and will be normal again in the next period. Stopping production costs €300,000 for sure. Stopping can also lead to not fulfilling a big order on time with a probability of 50% which would mean opportunity costs of €400,000.

(a) Draw a decision tree for the situation described.
(b) Which decision should Frieda Fear with a utility function $u(x) = 2 - 2x/10^6 - (x-10^6)/10^{12}$ make if she only looks at the costs for the next month?

9.13

An investor has constant absolute risk aversion. As you can see from the formula 9.29, the utility in case of uncertain final wealth which is normal distributed with expected value μ and variance σ^2 can also be defined as $\mu - 1/2\ c\ \sigma^2$.

The investor can choose between two alternative investment strategies:

– A riskless government bond, return 5%, riskless ($\sigma^2 = 0$);
– A risky bond with normal distributed return, expected return $\mu = 20\%$, variance $\sigma^2 = 0.25$.

The investor's risk parameter is $c = 0.1$, his original wealth w_0 is 100.

(a) First assume the investor wants to invest all of his wealth into only one of the two investment alternatives. Which of the two will he choose?
(b) Calculate the value of the risk parameter c (keeping all other parameters constant) which makes the investor indifferent between the two alternatives.
(c) After reading the bestseller *Live more happily by Diversification*, the investor realizes that he can achieve higher utility by investing in both alternatives. Calculate which split is optimal.
(d) Which characteristic of the normal distribution is especially precarious from a theoretical point of view concerning the description of possible returns of financial investment?

9.14

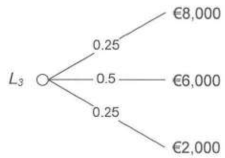
A expected utility maximizer has the utility function $u(x) = \alpha + \beta \log(x)$ with $\beta > 0$. Elicit at least 4 points on the utility function (including $u(x_{min})$ and $u(x_{max})$) such that it is normalized on the interval [€1,000, €11,000]. Apply the following method:

(a) Bisection version of the variable certainty equivalent method,
(b) Quantile version of the variable certainty equivalent method,
(c) Variable probability method,
(d) Method of equal utility differences,
(e) Trade-off method for utility functions; use x_a=€100 and x_b=€800.

9.15

Heinz-Lothar is a risk averse EUT decision maker. He makes the following preference statement regarding the following two lotteries.

(a) His risk premium (RP) for lottery L_1 is €1,000. What is his certainty equivalent (CE) for L_1?
What can be concluded for the certainty equivalent CE_2 that Heinz-Lothar assigns to lottery L_2? Provide an interval that contains all possible values of CE_2!
(b) Heinz-Lothar now compares L_1 to a third lottery L_3 with



Which one will he prefer, L_1 or L_3?
(c) How does the certainty equivalent for L_3 compare to the certainty equivalent for lottery L_2?

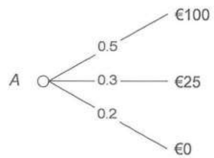
9.16

(a) Bill is expected utility maximizer. His risk premium for a lottery a = (€100, 40 %; €50, 40 %; €0, 20 %) is €10. What is his certainty equivalent (CE) for this lottery?
(b) Use the result from subtask 1 and the independence axiom to conclude what risk premium Bill assigns to the lottery b = (€100, 56%; €50, 16%; €0, 28%).

9.17

Angela is an expected utility maximizer with utility function $u(x) = \sqrt{x}$.

(a) Will Angela's decisions reflect risk aversion, risk neutrality, or risk proclivity?



(b) What is Angela's risk premium regarding the lottery A?

(c) Angela considers lottery A to be as attractive as a lottery B which has the consequences €100 and €0 only. What probabilities must be assigned to the outcomes of lottery B to result in such an indifference statement?

(d) Visualize the indifference statement above in a three-outcome-triangle.

(e) What is the slope of the indifference curves?

(f) Determine another lottery C that is in the interior of the same three-outcome-triangle (i.e. is not placed on the boundaries of the triangle) and has a certainty equivalent of €25. Do not use the explicit utility function to derive C but argue via the three-outcome-triangle.

Case Study: Petroleum and natural gas exploration at the Phillips Petroleum Company

Source: Walls et al. (1995).

In the late 1980s and early 1990s, the business division North American Eastern Onshore Exploration of the Phillips Petroleum Company was responsible for petroleum and gas explorations alongside the east and south coast. The managers wanted to use a consistent risk measure in the selection of the projects in order to compare projects with very different risks and to maintain a well-balanced portfolio of drilling projects. In addition, they wanted a technique to be able to decide on appropriate stakes in other companies' projects, to which they committed for the purpose of diversification.

The authors of the article developed a special computer program *Discovery*. Based on expected utility theory, it provided the opportunity to assess the projects in accordance with the management's risk preferences. The yardstick was the certainty equivalent of the project, which is the safe amount of money equivalent to the project fraught with risk according to the management's risk attitude.

The management of the business division settled for the assumption of constant risk aversion. This is (except for linear) only given for exponential utility functions. If the risk aversion coefficient c is known, the certainty equivalent can simply be calculated as:

$$CE = -\frac{1}{c}\ln\left(\sum_{i=1}^{n} p_i\, e^{-cx_i}\right)$$

where p_i is the probability of the consequence i and x_i is its monetary value.

For a given project, one could now examine the certainty equivalent for varous values of c. If one varied the shareholding quota in the project, one could compare the resulting functions and find out from which value of c on the certainty equivalent of a low, e.g. 50%, share exceeded the certainty equivalent of a higher, e.g. 100%, share. The managers could determine the optimal share in a given project for a fixed risk aversion measure c.

With the aid of the software it was possible to examine, how the risk of a project could be reduced by additional seismic information. The program computed the value of this additional information according to the expected profit as well as according to the certainty equivalent.

Since the investment capital was scarce, one strived to put the available drill projects in a preference order. For a fixed c, the program computes the shareholder quota with the highest certainty equivalent for each project and then arranges the options in descending order. This resulted in considerable differences in the ranking of the alternatives in comparison to the common criterion of expected profits.

In order to measure the risk attitude of the division management, the authors of the article analyzed the drilling decisions of the recent past. They found out that almost all decisions in the gulf coast region were consistent with risk aversion coefficients between $0.03 \cdot 10^{-6}$ and $0.05 \cdot 10^{-6}$. The management opted to take this level of c as a basis for the evaluation of new projects as well. From this point forward, all projects were analyzed with *Discovery*.

At the time of the article's formulation, a number of petroleum exploration firms used the software, sometimes only for occasional project decisions and sometimes to analyze the entire project portfolio.

Resistance of managers against the software on the one hand was based on a reluctance to quantify their risk aversion. On the other hand, every model is limited in its ability of representing reality. High demands concerning usability and transparency have to be made especially for the employment by managers without experience with decision theory; this is achieved by reducing the program's ability to model all possible complex cases.

Chapter 10: Decision under risk: incomplete information and multiple objectives

10.0 Summary

1. The expected value criterion can be extended to the case of incomplete information concerning the value function and probabilities.
2. There are two questions concerning decisions under incomplete information:
   - How can the incomplete information be modeled?
   - How can preference statements be obtained from incomplete information?
3. Preference statements can often be derived using simple linear programming (LP) approaches.
4. Risk analysis is an important method for decisions under incomplete information concerning the value function.
5. Sensitivity analysis is another possibility for supporting decisions under incomplete information.
6. In the case of multiple objectives, preferences can be modeled by multi-attribute utility functions.
7. The simplest form is the additive utility function, which is applicable if the requirement of the additive utility dependence of the attributes is fulfilled.
8. The condition for the mutual utility dependence is less restrictive. It implies a multiplicative utility function.

10.1 Model for decision under risk and incomplete information as well as sensitivity analysis

We have repeatedly pointed out that decision-makers do not necessarily have a precise preference. Consequently, for decisions under risk, the case of incomplete information and the sensitivity analysis also have to be taken into account, analogously to the approach in Section 6.5. A decision maker will, as an example, be able to state the approximate certainty equivalent of a lottery. If a method however requires the decision-maker to make accurate and complete preference statements, this requirement can lead to a rejection of the decision maker's decision propositions. Both sensitivity analysis and decisions with incomplete information afford the decision maker the possibility of obtaining results from less than accurate statements.

The *sensitivity analysis* starts by calculating the expected utility based on the exact value. The sensitivity of the decision with regard to the exact statement of the certainty equivalent is then analyzed, i.e. it is analyzed whether the decision changes if the certainty equivalent deviates within a predetermined range.

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_10, © Springer-Verlag Berlin Heidelberg 2010

Decisions under incomplete – or, synonymously, partial information – do not require the decision-maker to be able to state an exact probability distribution. It is sufficient that he can define a set of probability distributions which we will denoted by $P(I)$. $P(I)$ is the set of probability distributions that is compatible with the information set denoted by the letter I gathered by the decision maker. The decision maker also does not need a unique preference and hence does not need a unique value function u. The axiom of complete order is violated by the decision maker and the complete order is replaced by a partial one. The incompleteness of information concerning the utility function is modeled by the assumption that the decision maker has a set of utility functions denoted by $U(I)$. The expected utility criterion can be extended to the case of incomplete information as follows:

$$a \succeq b \Leftrightarrow \mathrm{EU}(a) \geq \mathrm{EU}(b) \quad \text{f. a. } p \in P(I) \text{ and } u \in U(I). \tag{10.1}$$

The formula states that an alternative a is preferred over an alternative b if the expected utility of a is greater than or equal to that of b for all feasible probability distributions and all feasible utility functions. If should be emphasized that the preference relation „$\succeq$" is defined with respect to the incomplete information sets $P(I)$ or $U(I)$; the preference can also be denoted „$\succeq_{P(I)}$" or „$\succeq_{U(I)}$".

The formula for the expected utility criterion in the case of incomplete information shows that the preference relation defined by the extended expected utility criterion does not need to be complete. The smaller the feasible sets of the probability distributions and utility functions are, the more complete the preference relation.

The extension of the expected utility criterion to the case of incomplete information can have great relevance in particular for the development of interactive procedures. If a decision maker wants to choose only between a small set of alternatives, a preference can often be deduced if the sets $U(I)$ and $P(I)$ are specified over rather small subsets. In some cases, it is sufficient for the class of feasible probability distributions that probabilities are specified for particular states. For the set of utility functions, it could be sufficient if the utilities are determined only for a small interval of consequences. This information can be elicited in such a way such that only the relevant intervals of the probability distribution and utility function are determined accurately.

In order for the application of the concept of incomplete information to be valid, two questions have to be answered in advance:

1. How can the sets $U(I)$ and $P(I)$ be determined and mathematically modeled?
2. Which algorithm can be applied for all utility functions and probability distributions to determine if the expected utility of an alternative is greater than that of a second alternative?

We will consider the sub-cases $P(I)$ and $U(I)$ separately in the following section. If the information about both the utility function and probability distribution is incomplete, there exist only few studies providing a possible solution to the questions above. We want to only mention these complex approaches here (see, for example, Pearman and Kmietowicz 1986 and Keppe and Weber 1989). In the following paragraphs of this chapter, we will first consider decisions under incomplete information followed by the sensitivity analysis.

In our opinion, the cases of incomplete information and sensitivity analysis are extremely relevant both practically and theoretically. In practice, decision makers are rarely willing or able to generate accurate information. For group decision-making, we also see a large set of possible applications as will be shown in Chapter 12. The case of incomplete information is interesting from a theoretical standpoint because it comprises the extreme case of complete probability information, i.e. traditional risk utility theory as well as the other extreme of complete ignorance. Ignorance refers to a case when no statements on probability are possible. The decision rules for the case of ignorance are presented in the literature. Textbook examples of decisions under ignorance are usually abstract decision matrices without practical interpretation; it is difficult to imagine situations in reality in which decision makers do not possess any idea about probabilities.

One example is that of an entrepreneur of a medium-sized company in Forst on the German Wine Route who has never conducted any business outside the Pfalz region watching a documentary about the South Seas Island Balla Balla on TV and spontaneously deciding to introduce his newly invented TV armchair to the island. Balla Balla has 1,000 inhabitants. The annual sales volume is unknown. Let us assume that no decision maker buys more than one armchair. With this assumption, there are 1,001 possible events, i.e. sales volumes of 0, 1, 2, …, 1,000. Even this uncertain situation is far from the theoretical concept of complete ignorance: the probability of selling an amount between 0 and 999, for example, is likely to be estimated higher by the entrepreneur than selling exactly 1,000 armchairs. Such a quantitative statement is, however, incompatible with the principle of ignorance.

As we are unconvinced by the relevance of the ignorance principle for business practice, the reader will not find a chapter on decisions under ignorance in this textbook. The entrepreneur in our example will at least have some intuition about the probabilities. Even if he does not have an accurate probability distribution for the events, he will be able to apply decision methods in the case of incomplete information to deduce a preference relation (for an overview of decisions under incomplete information, see Weber 1987).

10.2 Incomplete information concerning the probabilities *P(I)* or utility function *U(I)*

10.2.1 Incomplete information concerning the probabilities: *P(I)*

Given a limited set of n states of the world, the expected utility criterion for the case of incomplete information concerning the probabilities can be written as:

$$a \succeq b \Leftrightarrow \sum_{i=1}^{n} p_i u(a_i) \geq \sum_{i=1}^{n} p_i u(b_i) \qquad (10.2)$$

$$\text{f. a. } p \in P(I).$$

The set of feasible probability distributions can be defined by simple restrictions. Usually, decision makers would be able to state that the probability of a specific state of the world is above or below a particular value or that the probability of an event lies within a particular interval. Apart from allocating incomplete probability statements to a particular state of the world, probabilities can also be compared to each other. The probability of one event, for example, can be estimated to be higher than the probability of a second event. The possibilities of modeling incomplete information mentioned above – stating probability intervals and the ordinal comparison of probabilities of two states – can be formally written as:

$$p_i^- \leq p_i \leq p_i^+ \qquad (10.3)$$

$$p_i \leq p_j \qquad (10.4)$$

There are, of course, other possible approaches to modeling the set of feasible probability distributions. However, these two approaches have the advantage that this information is easily achievable from decision makers and that the expected utility values can be easily calculated for this probability information. The question of whether the expected value of one alternative is greater than the expected value of a second alternative for all feasible distributions is formally similar to the decision problem under incomplete information concerning the objective weights (Section 6.5.3). Therefore, the following linear programming problem has to be solved for the pair of alternatives a and b:

$$\begin{aligned} &\text{Maximize or minimize} \sum_{i=1}^{n} p_i \left[u(a_i) - u(b_i)\right] \\ &p_i^- \leq p_i \leq p_i^+ \\ &p_i \leq p_j \\ &\sum_{i=1}^{n} p_i = 1 \\ &p_i \geq 0. \end{aligned} \qquad (10.5)$$

If the minimum – and hence also the maximum – of the objective function is greater than zero, alternative a is preferred over alternative b. If the maximum – and thus also the minimum – is less than zero, alternative b is preferred over alternative a. If the maximum is greater and the minimum is smaller than zero, no clear preference statement for a or b with regard to the class of probability distributions defined by the restrictions can be made so far. In this case, the decision maker has to provide additional information to reduce the set of feasible probability distributions. He could decrease the length of the intervals possible or state additional restrictions. The restrictions (i.e. modeling the feasible set of probability distributions) can be formulated in arbitrarily complex ways. After being given new information, the dominance can again be tested using the LP approach. The process has to be continued until the desired decision can be made.

The model presented here makes use of the LP approach. If there are only restrictions according to (10.3), the minima and maxima can also be determined with simpler algorithms (Sarin 1978). We will illustrate this procedure with an example shortly. Further work on decisions under incomplete information concerning the probabilities can also be seen in Brachinger (1982) and Rios Insua (1990).

Decisions under incomplete information can nowadays be supported efficiently by computers. Linear programs can be solved rapidly and automatically by standard software packages. The approaches for the practical cases presented here can be applied even more easily and efficiently with particular decision analysis software packages that support decisions under incomplete information.

Let us exemplify the case for a decision under incomplete information concerning probabilities. Table 10-1 shows a decision matrix with alternatives a, b, and c, states of the world s_1, s_2, and s_3 and the alternatives' utilities for the different states of the world.

Table 10-1: Example of a decision with incomplete probability information

	s_1	s_2	s_3
a	0.375	0.125	0.25
b	0.25	0.5	0.19
c	0.28	0.25	1

$$0.5 \le p(s_1) \le 0.7 \qquad 0.1 \le p(s_2) \le 0.3 \qquad 0.1 \le p(s_3) \le 0.3$$

Comparing alternatives a and b leads to the following LP approach:

$$\text{Min} \quad p(s_1)\cdot(0.375-0.25)+p(s_2)\cdot(0.125-0.5)+p(s_3)\cdot(0.25-0.19)$$

$$\begin{aligned} p(s_1) &\ge 0.5 \\ p(s_1) &\le 0.7 \\ p(s_2) &\ge 0.1 \\ p(s_2) &\le 0.3 \\ p(s_3) &\ge 0.1 \\ p(s_3) &\le 0.3. \end{aligned}$$

The LP approach results in:

$$\begin{aligned} &\text{Min} \quad [\mathrm{EU}(a)-\mathrm{EU}(b)]<0 \quad \Rightarrow a \text{ does not dominate } b, \\ &\text{Max} \quad [\mathrm{EU}(a)-\mathrm{EU}(b)]>0 \quad \Rightarrow b \text{ does not dominate } a, \end{aligned}$$

$$\begin{aligned} &\text{Min} \quad [\mathrm{EU}(a)-\mathrm{EU}(c)]<0 \quad \Rightarrow a \text{ does not dominate } c, \\ &\text{Max} \quad [\mathrm{EU}(a)-\mathrm{EU}(c)]<0 \quad \Rightarrow c \text{ dominates } a. \end{aligned}$$

Comparing alternatives b and c, we would like to illustrate how you can test for dominance manually in a case in which there are only constraints of the form (10.3). To begin with, you should reduce your workload and limit your test for dominance to one possible direction. You can do this by choosing an arbitrary probability combination (e. g. $p(s_1)=0.6$; $p(s_2)=0.2$ and $p(s_3)=0.2$) and determining that c has a higher expected utility in this case. Hence you can confine yourself to analyze whether b is dominated by c.

For this purpose, we have to maximize the term EU(b)–EU(c). This term can be written as:

$$p(s_1)\cdot(0.25-0.28)+p(s_2)\cdot(0.5-0.25)+p(s_3)\cdot(0.19-1)$$

or

$$p(s_1)\cdot(-0.03)+p(s_2)\cdot(0.25)+p(s_3)\cdot(-0.81).$$

To maximize this term, we initially set all probabilities to their lower limit and increase the single probabilities successively (up to their upper limit) in such a way that the positive effect becomes as large as possible (we actually want to maximize this difference). In this example, the three probabilities would initially be set to their minimum value

$$p(s_1)=0.5; \quad p(s_2)=0.1 \quad \text{and} \quad p(s_3)=0.1$$

and using the remaining probability mass of 0.3, we would increase $p(s_1)$ from 0.1 to its maximum value of 0.3 (note that state 2 has the largest utility difference, namely +0.25). We then increase the probability for state 1 to 0.6 using the residual probability mass of 0.1 because this state has the second largest utility difference (as the total probability must be increased to 1 even if this lowers the utility difference). The maximum value of the expected utility difference is thus achieved for $p(s_1)=0{,}6$, $p(s_2)=0{,}3$ und $p(s_3)=0{,}1$ and amounts to $-0{,}024$. If the maximum value is negative, alternative b is dominated by c; alternative c is the optimal alternative.

10.2.2 Incomplete information concerning the utility function $U(I)$

In the case of incomplete information concerning the utility function we assume that the probability information can be exactly determined. One alternative is preferred over a second alternative if the expected utility of the first alternative is greater than the expected utility of the second alternative for all feasible utility functions. This is formally written as:

$$a \succeq b \Leftrightarrow \sum_{i=1}^{n} p_i\, u(a_i) \ge \sum_{i=1}^{n} p_i\, u(b_i) \text{ f. a. } u \in U(I). \qquad (10.6)$$

A class of feasible utility functions can be determined by the approaches to identify utility functions presented in Chapter 9. Some approaches can be directly applied in the case of incomplete information (for example, the quantile method) whereas for some other approaches the application is slightly more complex (for example, for the midpoint chaining technique). When applying such an approach under incomplete information, the exact utility function will be replaced by a spectrum of utility functions that are compatible with the information gathered in the determination process. Let us exemplify the approach using the quantile version of the variable certainty equivalent method.

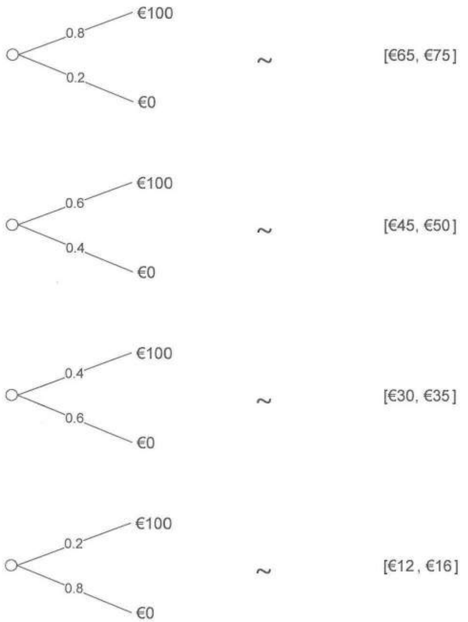


Figure 10-1: Intervals for certainty equivalents using the quantile method

Let us assume that the decision maker wants to determine the utility function for the interval [€0, €100]. He will be presented with four lotteries for which he will state the intervals of certainty equivalents. Figure 10-1 depicts the four lotteries and the elicited certainty equivalents. As an example, the figure shows that the decision maker evaluates the lottery (€100, 0.6; €0, 0.4) with a certainty equivalent interval of [€45, €50].

The class of compatible utility functions can be directly determined from the data, as shown in Figure 10-2.

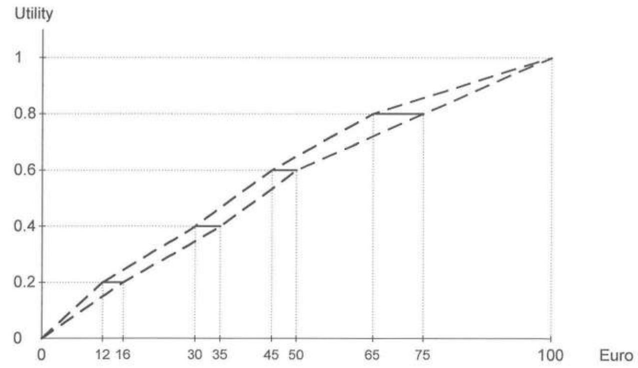


Figure 10-2: Class of feasible utility functions

The class of utility functions was determined by replacing the nodes of an exact risk utility function by "interval nodes". Using the particular construction of the basis reference lottery, we can deduce that the money amounts yielding a utility of 0.6 are within the interval [€45, €50]. Analogously to the procedure for value functions and utility functions, the end points of the ranges can be linearly interpolated. In this example, it would also be possible to draw a function of predefined functional form through the left or right end points. (Nearly) all possible, i.e. feasible, monotonically increasing, functions are within the area indicated in Figure 10-2.

In a follow-up step one can verify whether it is possible to make preference statements on the basis of the determined class of utility functions for given alternatives. If the ranges of the utility functions are defined by linear functions (as in Figure 10-2) preference statements can be verified by using a simple linear programming approach. The values $u(a_i)^-$ and $u(a_i)^+$ indicate the minimum and maximum values for the utility of the consequence a_i.

$$\text{Maximize or minimize} \sum_{i=1}^{n} p_i \left[u(a_i) - u(b_i) \right] \tag{10.7}$$

$$u(a_i)^- \le u(a_i) \le u(a_i)^+$$

$$u(b_i)^- \le u(b_i) \le u(b_i)^+.$$

Such a check for dominance represents a very tough test. Note that when formally maximizing the expected utility difference it is admissible that, in the case of identical consequences $a_i = b_i$, a_i assumes it upper limit $u(a_i)^+$ whereas b_i is set to its lower limit $u(b_i)^-$. This is certainly not what is usually meant by "uncertainty about the utility function". It would not be very convincing if a dominance relation were to fail due to such artifacts. For these cases, milder dominance tests can be defined that pose additional, "more reasonable" requirements regarding the valuation of consequences (e.g. that identical consequences have to be evaluated identically or that the chosen utility valuations must increase monotonically with the consequence levels).

Our approach for describing the set of feasible utility functions follows the approach for describing the set of feasible probability distributions. There is however a second and widely used way to describe the utility function $U(I)$. In this approach, the set of feasible utility functions is characterized in such a way that all feasible utility functions possess particular properties. These properties are, for example, that the utility functions are monotonically increasing or decreasing, or that the decision maker is risk averse and his utility function is concave. The class of strictly monotonically increasing utility functions

$$U_M(I) = \{u \mid x_i > x_j \Rightarrow u(x_i) > u(x_j)\} \tag{10.8}$$

is of particular importance for many economic applications. The set $U_M(I)$ only comprises the preference information that more (of the relevant value) is better than less. This preference concept (for the case of incomplete information) which is based on a defined set of utility functions is of great importance in the areas of investment and finance theory. After specifying the feasible class of utility functions, the preference relation is verified:

$$a \succeq b \Leftrightarrow \sum_{i=1}^{n} p_i\, u(a_i) \geq \sum_{i=1}^{n} p_i\, u(b_i), \quad u \in U_M(I). \tag{10.9}$$

A simple procedure is needed that determines, for any given alternatives a and b whether a is better than b for all permissible utility functions.

Fortunately, the preference can be directly checked using the distribution functions of the alternatives a and b. For this purpose, we have to introduce the concept of stochastic dominance.

Definition 10.1 (stochastic dominance)
An alternative a stochastically dominates alternative b if for each consequence of the objective variable, the probability of exceeding this consequence is at least as high for a as for b and for at least one consequence of the objective variable the probability of exceeding this consequence is higher for a than for b.

If an alternative a stochastically dominates an alternative b, it can be shown that for all strictly monotonically increasing utility functions, the expected utility of a is higher than of b.

To determine stochastic dominance, you have to determine the distribution function $P_a(x)$ of the target variable X for each alternative a. This distribution function indicates the probability of the target value x being achieved or surpassed for alternative a. In Chapter 8, we have shown how complex distribution functions can be generated by simulation. $1 - P_a(x)$ can then be defined as a function indicat-

ing the probability that alternative a will exceed the value x. Mathematically, the concept of stochastic dominance can be described in the following way: a dominates b stochastically if and only if $1 - P_a(x) \geq 1 - P_b(x)$ for all values of X and $1 - P_a(x) > 1 - P_b(x)$ for at least one value of X.

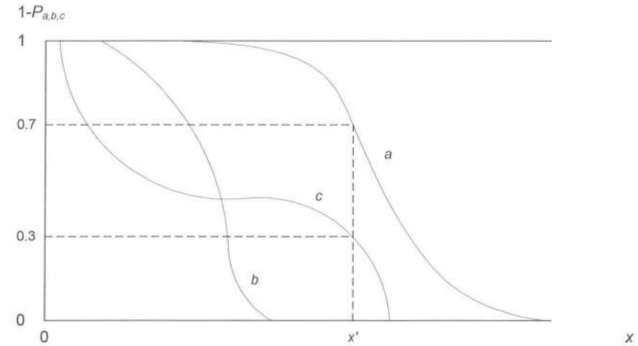


Figure 10-3: Risk profile for the three alternatives a, b, and c

The concept of stochastic dominance becomes particularly transparent if the functions $1 - P_a(x)$ are illustrated graphically. In this case – as described in Chapter 8 – the functions are called risk profiles. The analysis of risk profiles is often referred to as *risk analysis*.

Figure 10-3 shows the risk profiles of the three alternatives a, b, and c, highlighting the fact that alternative a stochastically dominates both alternatives b and c. For each possible consequence x, a has a higher probability of exceeding this value. For example, a exceeds value x' with a probability of 0.7, c exceeds value x' with a probability of 0.3 and b stays below value x' with certainty. Stochastic dominance cannot provide support for any decision between b and c. If risk profiles intersect each other, the decision maker must provide more information about his value function, meaning that he must reduce the quantity of permissible value functions $U(I)$. To this end, higher order concepts of stochastic dominance were developed for making more preference statements. For second order stochastic dominance, for instance, the utility function must not only be monotonically increasing but also concave. To gain an overview of more stochastic dominance concepts and research results, see Bawa (1982) and Levy (1992).

The concept of stochastic dominance can also be applied if the decision maker additionally only has incomplete information about his probabilities. The risk profiles are replaced by bandwidths of profiles that can be evaluated by analogous concepts of stochastic dominance (Keppe and Weber 1993).

In this chapter, we have embedded the concept of stochastic dominance into the context of the expected utility theory. Against the background of this theory, proposals for decisions between intersecting risk profiles must be assessed. A decision maker might for instance prefer the alternative with the lower probability of loss (shortfall probability). This procedure tries to model the risk aversion of decision makers. Such proposals are usually incompatible with risk utility theory and do not define a preference, which in turn would be characterized by a convincing system of axioms; we will therefore not discuss such proposals.

The expected utility criterion defines a complete order of the set of alternatives. The stochastic dominance criterion needs less preference information and thereby cannot define as many preferences between the alternatives.

An alternative which is optimal ex ante needs not to be optimal after the uncertainty is resolved (ex post). Let us consider, for instance, the decision matrix in Table 10-2. In this example, alternative b stochastically dominates alternative a. Nonetheless, alternative a can be better than b *ex post*. To exclude this possibility, an even stronger dominance requirement would have to be fulfilled, namely state-wise dominance. This dominance criterion was already briefly discussed in Section 1.2.6; state-wise dominance is present if an alternative a has a higher utility consequence than alternative b in each possible state of the world. In this case, the decision maker has no reason to regret his choice ex post as he will always be better of with alternative a. State-wise dominance always implies stochastic dominance but not the other way around. However, this relation does not apply vice versa, as the example of Table 10-2 shows. For the expected utility criterion under both complete and incomplete information, the alternative optimal ex ante can lead to a suboptimal outcome ex post. However, this should not deter the decision maker from applying the expected utility criterion or its extension to incomplete information. Of course, hindsight is 20/20 so it is important to make a good decision in advance, i.e. a decision based on an axiomatically risk/utility calculus.

Table 10-2: Stochastic dominance of b

	s_1 $p(s_1)=0.7$	s_2 $p(s_2)=0.3$
a	2	3
b	3	2

The general approach of a risk analysis is illustrated in the following example. A farmer can cultivate either sweet corn, wheat, or hops on his field, i.e. $A = \{\text{sweet corn, wheat, hops}\}$. Three states are possible regarding the weather: $S = \{\text{dry, average, humid}\}$. The decision matrix in Table 10-3 depicts the probabilities and performance consequences (in €) for this example. Any experts in the agricultural business will most likely need to excuse our numbers.

Table 10-3: Example of a risk analysis

	Dry	Average	Humid
Probability	0.2	0.5	0.3
Sweet corn	2,500	5,000	3,500
Wheat	3,000	2,500	2,000
Hop	1,800	3,500	1,800

If you deduce from the decision matrix probability functions P and subsequently the functions $1-P_{Corn}(x)$, $1-P_{Wheat}(x)$, and $1-P_{Hop}(x)$, you obtain the data in Table 10-4.

Table 10-4: Risk profiles of the three alternatives

X	$1-P_{Corn}$	$1-P_{Wheat}$	$1-P_{Hop}$
1,800	1	1	0.5
2,000	1	0.7	0.5
2,500	0.8	0.2	0.5
3,000	0.8	0	0.5
3,200	0.5	0	0.5
3,500	0.5	0	0
5,000	0	0	0

The risk profiles for the example are depicted in Figure 10-4. The alternative sweet corn stochastically dominates the cultivation of both wheat and hops. A preference between wheat and hops cannot be derived using the criterion of stochastic dominance. If the decision maker wanted to determine his second best alternative, he would have to specify his utility function by indicating additional general characteristics or with the help of concrete indifference statements.

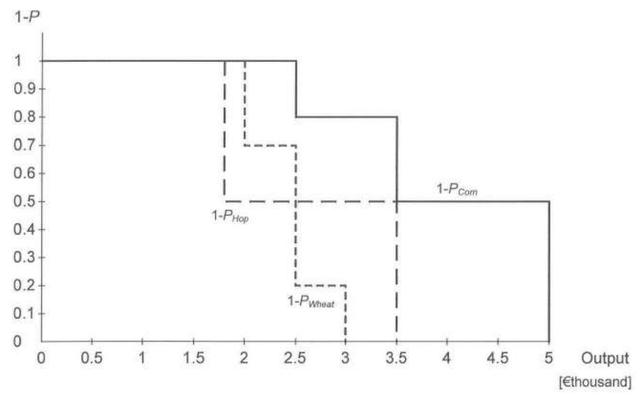


Figure 10-4: Risk profile for the cultivation example

10.3 Sensitivity analyses

A sensitivity analysis generally examines how an objective value or decision depends on the respective relevant parameters. We have introduced sensitivity analyses in Section 6.5.4 when discussing incomplete information regarding objective weights.

Decisions under incomplete information with respect to probabilities and utility functions are closely related to the concept of the sensitivity analysis. In the latter concept, data is usually elicited on a point value basis; the variation interval in which the solution remains stable is subsequently analyzed. In the approach with partial information, we abstain from information with exaggerated (i.e. spurious) precision and try to find a solution using partial but reliable information. For decisions under incomplete information, it can be shown how decisions depend on probability and utility judgments. Critical values can be defined for which two alternatives are equally good. As we have addressed before, the analysis of critical values within the scope of decision making under incomplete information can particularly be used for the control of an interactive decision support system.

In some cases, a sensitivity analysis under incomplete information can also provide graphical decision support. This is particularly useful in the presence of incomplete probability information for three or fewer states. The approach and explanatory power of this kind of sensitivity analysis is best explained using an example:

A decision maker decides between three alternatives *a*, *b*, and *c* whose consequences depend on the states s_1 and s_2. The probability that s_1 occurs is denoted *p*, the probability of s_2 is $1 - p$. Given the consequences and utility functions, the ex-

pected utilities of all three alternatives only depend on the probability *p*. The decision matrix in Table 10-5 indicates the utility of the consequences depending on the respective states.

Table 10-5: Example of a sensitivity analysis with respect to *p*

	s_1	s_2
	p	$1-p$
a	0.66	0.5
b	0.5	0.83
c	0.25	1

The alternatives have the following expected utility values:

$$\mathrm{EU}(a) = 0.5 + 0{,}16 \cdot p$$
$$\mathrm{EU}(b) = 0.83 - 0{,}33 \cdot p$$
$$\mathrm{EU}(c) = 1 - 0.75 \cdot p$$

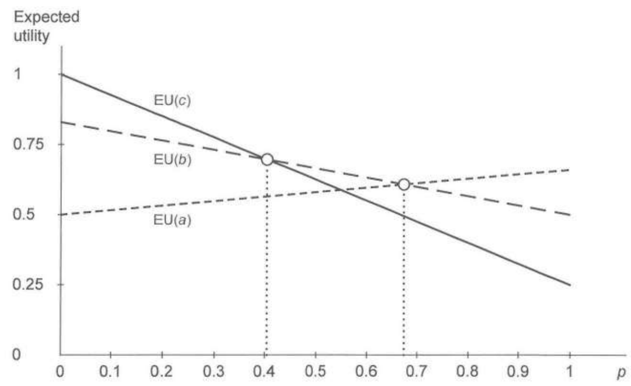


Figure 10-5: Sensitivity analysis for probability *p*

To determine the optimal alternative, it is not necessary to specify precisely the probability *p*; in Figure 10-5, the expected utilities of the alternatives are depicted, depending on the probability *p*. If the decision maker, for instance, thinks that *p* is at least 0.75, alternative *a* is optimal. If he indicates the interval [0.1; 0.3] for *p*, he will pick alternative *c*. The critical probability values for this sensitivity analysis are $p = 0.4$ und $p = 0.67$.

10.4 Decision making under multiple objectives

10.4.1 The additive model

In this section, we will discuss decisions for which there are multiple objectives and the consequences of alternatives are risky. In an extension to the ideas of the multiattribute utility function (Chapter 6) and utility function for one objective (Chapter 9), we will introduce a multiattribute utility function. Such functions should represent the decision maker's preferences of the different attributes and risk.

We will use the following notation: in the case of known (i.e. sure) consequences, the alternatives $a \in A$ have been characterized by vectors of the form $a = (a_1, ..., a_m)$. a_r represents the attribute level of an alternative a relating to attribute X_r. In the case of risk, the alternative is determined by a distribution of possible attribute level combinations. We extend the notation to write

$$a = [p_1, (a_{11}, ..., a_{1m});\ p_2, (a_{21}, ..., a_{2m});\ ...;\ p_n, (a_{n1}, ..., a_{nm})]$$

for an alternative a that leads to a consequence $(a_{i1}, ..., a_{im})$ with a probability p_i.

In other words, if the decision maker chooses alternative a, there are n possible consequences. The condition or event i occurs with a probability p_i and leads to the consequence $(a_{i1}, ..., a_{im})$; a_{ir} represents the values of the objective variables r if alternative a is chosen and state i occurs. Table 10-6 illustrates the attribute levels of an alternative a with their respective probabilities.

Table 10-6: Levels a_{ir} of an alternative a relating to attribute r in state i

	States	s_1	...	s_i	...	s_n
	Probabilities	p_1	...	p_i	...	p_n
	1	a_{11}	...	a_{i1}	...	a_{n1}
	...	...	...	...	...	...
Attribute	r	a_{1r}	...	a_{ir}	...	a_{nr}
	...	...	...	...	...	...
	m	a_{1m}	...	a_{im}	...	a_{nm}

We model the decision maker's preferences using a multiattribute utility function $u(x_1, x_2, ..., x_m)$ that indicates for (any) two alternatives a and b:

$$a \succeq b \Leftrightarrow \mathrm{EU}(a) > \mathrm{EU}(b). \tag{10.10}$$

In other words, a is preferred over b if the expected value of the utility of a is greater than the expected value of the utility of b.

In the sense of the decomposition principle, it would be desirable that the utility function can be decomposed in an additive way. Let $k_r > 0$ with $\Sigma k_r = 1$ be the attribute weights. An additive utility function is then defined by

$$u(x_1, x_2, ..., x_m) = \sum_{r=1}^{m} k_r \cdot u_r(x_r), \tag{10.11}$$

with $u_r(x_r)$ representing the one-dimensional utility function (normalized to be between zero and one) for the attribute X_r.

For an additive utility function the expected value of utility of the risky alternative a is

$$\mathrm{EU}(a) = \sum_{i=1}^{n} p_i \cdot \left[\sum_{r=1}^{m} k_r u_r(a_{ir})\right] \tag{10.12}$$

or, equivalently

$$\mathrm{EU}(a) = \sum_{r=1}^{m} k_r \left[\sum_{i=1}^{n} p_i u_r(a_{ir})\right]. \tag{10.13}$$

According to the first formula, the utility of an alternative equals the sum of the probability weighted utilities of possible consequences; according to the second formula, however, the utility of an alternative equals the sum of the expected values (multiplied by the attribute weights) of the single attribute utilities.

The one-dimensional utility functions u_r are determined as outlined in Chapter 9. The weighting factors k_r can be determined using the trade-off method analogously to the procedure for multiattribute value functions as illustrated in Chapter 6.

The swing method cannot be applied for the objective weighting of utility functions. This method requires the decision maker to specify preference differences, for instance, the extent to which the preference difference between alternatives a^- and b^1 is greater than that between a^- and b^2. Such statements, however, cannot be obtained from utility functions because these do not measure value differences.

10.4.2 Condition for the additive model: additive utility independence

To model preferences with an additive utility function, they have to fulfill the condition of additive utility independence.

Definition 10.2 (Additive utility independence)
The attribute set $X_1, ..., X_m$ is called *additively utility independent* if the preferences over lotteries depend only on the distributions of single attribute levels but not on the distributions of attribute combinations.

Let us start with an example. A tourist planning a weekend trip to the mountains considers there to be a 50% chance of snow and also a 50% chance of sunshine. The probabilities are valid for both winter sports resorts a and b; however, the alternatives are different with respect to the joint probabilities of events. For a there is either snow or sun, whereas for b there is either both or nothing, as illustrated in Table 10-7.

Table 10-7: Additive utility independence implies indifference between a and b

	0.5	0.5
a	Sun, no snow	Snow, no sun
b	Sun and snow	Neither sun nor snow

In both cases the (marginal) probabilities of the attributes are equal. You hence need to be indifferent between both alternatives for the additive utility model to be applicable. This can be seen directly from the expected utility formula, which is independent of the probabilities of attribute level combinations.

In many situations, the condition for additive utility independence is *not* fulfilled. Many decision makers prefer alternative a because they would at least "receive something". However, there will be some who want "all or nothing" and prefer the more extreme alternative b.

Additive utility independence mainly fails in the following cases (von Winterfeldt and Edwards (1986), p. 346 ff.): the first case is the presence of substitutive or complementary relationships among attributes. Some people enjoy skiing more when the sun is shining; snow and sunshine have complementary characteristics. Those people would prefer alternative b. For others, the attributes have a substitutive characteristic. If they go skiing sunshine is not important, but if there is no snow they appreciate sunshine for other recreational activities. Those people would prefer alternative a.

"Intrinsic" multiattributive risk aversion (or seeking) represents a second case of missing additive utility independence. Assume you like snow to go skiing and you like sunshine but your pleasure derived from going skiing is independent of sunshine, and your pleasure derived from sunshine is independent of snow. If you prefer a over b you are multiattributively risk averse; for the case $a \succ b$ you are multiattributively risk seeking with regard to both attributes. In this case, risk aversion must be interpreted in the sense that the decision maker seeks to avoid the risk of extremely disadvantageous consequences and – for given marginal probabilities – prefers alternatives with only moderate consequences.

To determine whether your preferences fulfill the condition of additive independence in a particular practical case, you should compare hypothetical lotteries for all attribute pairs for which

(a) There is a 50% probability that attribute 1 attains the highest and attribute 2 attains the lowest level; otherwise, attribute 1 attains the lowest and attribute 2 attains the highest level,
(b) There is a 50% probability that both attributes attain their respective highest level; otherwise, both attributes attain their respective lowest level.

The additive model is suitable if you are (almost) indifferent for all comparisons. The lower the difference between the best and worst attributes, the more likely the condition of the additive utility independence will be fulfilled.

As the additive utility model can be managed the most easily of all models, it is advisable to attempt to model the situation in such a way that its assumptions are fulfilled, if this is not already the case from the start. If the missing independence is because of complementary or substitutive relationships among attributes, you should remove these relationships by redefining the objectives. We have already discussed this point when dealing with the additive *value function*. Let us consider, for example, the attributes "reliability" and "service quality" for the choice among different car brands (Clemen (1996), p. 585). You are unsure about car brands. You neither know the cars' reliability nor the performance of the corresponding garages. There is some substitutability between the two attributes; if the car rarely breaks down, service quality is less important. Consequently, you will not find additive utility independence a useful concept and you will probably prefer a "middle" option (either the car is reliable or the garages are good, each with a probability of 50%), compared to an "all or nothing" option (i.e. a 50% probability that both the car is reliable and the garage is good). You could eliminate the interdependence between the attributes by focusing on more fundamental objectives. Ultimately, you might wish to minimize the downtime of your car or repair costs; in this case, you remove reliability and quality of service and replace them with the actual goal.

If there is mutual preference independence within the attribute set so that you can determine an additive value function (Chapter 6), there are no substitutive or complementary interdependencies, and the chances that the additive utility model holds are good. If there is still no additive utility independence, this might be due to a multiattribute "intrinsic" risk aversion or seeking (as already mentioned). The multiattribute "intrinsic" risk aversion stems from the same psychological causes as the risk attitude for one-dimensional utility functions. The risk averse decision maker tends to prefer "mediocre" lotteries instead of those with extreme consequences. In this case, non-additive utility models are available (Keeney and Raiffa 1976, von Winterfeldt and Edwards 1986, French 1988), of which we will only outline the multiplicative model below.

10.4.3 The multiplicative model

A multiplicative utility function $u(x) = u(x_1, x_2, ..., x_m)$ is defined by (see Keeney and Raiffa 1976, p. 289 and p. 325)

$$u(x) = \frac{\prod_{r=1}^{m} [kk_r u_r(x_r) + 1] - 1}{k} \tag{10.14}$$

where u_r is a one-dimensional utility function (normalized to be between zero and one), k_r are scaling constants and k is calculated as

$$1 + k = \prod_{r=1}^{m} (1 + kk_r). \tag{10.15}$$

The form (10.14) might appear usually complicated, particularly because of the recursively defined factor k. It has, however, only a normalizing role and guarantees that the function $u(x)$ remains between 0 and 1.

For example, for $m = 2$ attributes, it holds for the utility of consequence $x = (x_1, x_2)$ under (10.14) that

$$\begin{aligned} 1 + ku(x) &= (kk_1u_1 + 1) \cdot (kk_2u_2 + 1) \\ u(x) &= k_1u_1 + k_2u_2 + kk_1k_2u_1u_2 \end{aligned} \tag{10.16}$$

with

$$k = \frac{1 - k_1 - k_2}{k_1k_2}.$$

The utility of x consists of the weighted individual utilities of the attributes and a multiplicative term which reflects the interaction of attributes.

The following three cases are possible:

$$\begin{aligned} &\sum k_r = 1, \quad k = 0 \\ &\sum k_r < 1, \quad k > 0 \\ &\sum k_r > 1, \quad -1 < k < 0. \end{aligned} \tag{10.17}$$

In the first case, one obtains the additive model; there are no utility interactions among the variables. In the other cases, the constant k uniquely determines the form of interactions. In the second case, all interactions are complementary, and in the third case there are complementary and substitutive interactions depending on the exponent k (von Nitzsch 1992, p. 61 ff.).

For the multiplicative utility model, mutual utility independence is required; this is a weaker condition than additive utility independence. We will define mutual utility independence in the next section.

If this condition is fulfilled and you want to apply the multiplicative model, you proceed as follows:

1. Determine the individual utility functions over the attributes,
2. Determine the weighting factors k_r as in the additive model,
3. Determine the interaction constant k.

The third step is iterative. If the sum of k_r equals 1, the additive model has been obtained and you are done. If $\Sigma\, k_r > 1$, it must hold that $-1 < k < 0$. To determine k iteratively, you assume tentatively a negative value k' and test if

$$1 + k' = \prod_{r=1}^{m} (1 + k'k_r) \tag{10.18}$$

holds true. If the left side is greater, it must hold that $k < k'$, i.e. you must reduce k'. If the right side is greater, it holds that $k > k'$, and you must try a higher value.

This procedure is repeated until Equation (10.19) holds. This calculation can be much simplified using a spreadsheet program (e.g. goal seek in Excel).

There is an alternative method. First, you determine the additive multiattribute *value* function $v(x) = \Sigma w_r\ \ v_r(x_r)$, as described in Chapter 6. If the number of attributes m is at least 3, it holds that there is a unique relationship between the additive value and multiplicative utility function (Keeney and Raiffa (1976), pp. 330 ff.). The utility function normalized between zero and one takes one of the three following forms:

$$\begin{aligned} &\text{(a)}\quad u(x) = \frac{1 - e^{cv(x)}}{1 - e^{c}} \quad \text{with} c < 0 \\ &\text{(b)}\quad u(x) = v(x) \\ &\text{(c)}\quad u(x) = \frac{1 - e^{cv(x)}}{1 - e^{c}} \quad \text{with} c > 0. \end{aligned} \tag{10.19}$$

After determining the value function $v(x)$, you only need to find the parameter c of the exponential function. For this purpose you have to determine one point on the utility function; you could do this as follows: you take an individual utility function over an arbitrary attribute, let's say $u_r(x_r)$, and estimate the certainty equivalent x_r* of a lottery with equal chances of the lowest and highest attribute level. This certainty equivalent has a utility of $k_r/2$. If the value of the certainty equivalent equals its utility, i.e. if $v_r(x_r^*) = k_r / 2$, the additive model (b) has been obtained and $u(x) = v(x)$. For $v_r(x_r^*) < k_r / 2$ form (a) holds and for $v_r(x_r^*) > k_r / 2$ form (c) holds. In both cases, you insert $u(x) = k_r / 2$ and $v(x) = v_r(x_r^*) \cdot k_r$ into the appropriate exponential function and then solve for c.

The utility function according to (10.14) has to be the same as that according to (10.19). You can hence apply both methods in parallel to test the consistency of both measurements.

10.4.4 Condition for the multiplicative model: mutual utility independence

This condition is the analogue to the mutual preference independence in the case of no risk.

Definition 10.3 (Mutual utility independence)
An attribute X_r is utility independent of the other attributes if the preferences over lotteries which are only different in the levels of X_r are independent of the fixed levels of the other attributes. *Mutual utility independence* means that each subset of attributes is utility independent of its complementary set.

To test whether there is mutual utility independence, analyzing all possible 2^m – 2 subsets is unnecessary. There are weaker conditions that are sufficient to guarantee mutual utility independence (Keeney and Raiffa 1976, p. 292). It is sufficient, for example, that each subset consisting of $m - 1$ attributes is utility independent of the residual attribute. You hence have to pick each attribute individually and ask yourself the following question: does the level of this attribute have a consequence for the preferences regarding lotteries for which the consequences regarding the other attributes differ from each other?

The procedure is principally the same as testing preference independence; however, it asks for preferences over lotteries instead of preferences over certain consequences.

Questions and exercises

10.1

A decision alternative a has the following possible financial consequences (costs in €) and probabilities stated in the table below.

20,000	22,000	24,000	26,000
10%	25%	35%	30%

A second alternative b can have the consequences €20,000 or €24,000 both of which have a probability of 50%.

(a) Draw the risk profiles of a and b.
(b) Can you decide on one alternative?

10.2

Next to the two known dominance concepts of stochastic dominance and state-wise dominance we would like to introduce a third concept, namely the one of absolute dominance. Absolute dominance is present if the worst consequence of an alternative is still better than the best consequence of another alternative.

(a) Give an example for the two actions a and b for which a dominates b stochastically but not absolutely.
(b) From looking only at the risk profiles, can you determine absolute dominance?

10.3

A company has to choose among three investments a, b, and c. The net present values of the projects depend on the economical development and are shown in the table as three different scenarios (in millions €).

	s_1	s_2	s_3
a	10	15	5
b	5	5	20
c	2	10	15

There exist only rough specifications about the probabilities: $0.1 \leq p(s_1) \leq 0.3$, $0.2 \leq p(s_2) \leq 0.3$, and $0.3 \leq p(s_3) \leq 0.5$.

The company aims for maximal net present value. Can the decision already be made with the intervals given for the probabilities?

10.4

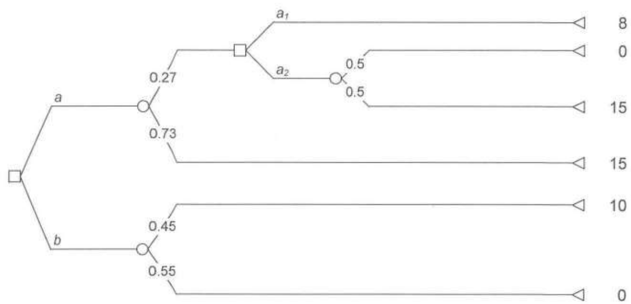
The speculator Sam has to chose among three investment alternatives which lead to the wealth positions stated in the table (in millions €), depending on the state of the world. The only thing Sam knows about his utility function is that it increases with increasing wealth.

	s_1	s_2	s_3
a	4	2	1
b	1	2	5
c	4	2	5

(a) Which preference order will a rational (in terms of risk utility theory) speculator chose for the actions a, b, and c?
(b) Assume alternative c is not available anymore. Sam picks b and state s_1 occurs. A friend accuses Sam of having decided wrongly because a would have resulted in a higher wealth position. What do you think about this accusation?

10.5

Calculate the risk profiles of the three possible strategies from the following decision tree. Is one strategy stochastically dominated?



10.6

A manufacturer with the utility function $u(x) = \sqrt{x}/30$ has to decide between three investment alternatives (a, b, c). The possible results (net present values in thousands €) depend on whether a particular important supplier is bought by a competitor or not.

	Supplier not bought	Supplier bought
a	818	676
B	676	900
C	784	784

Perform a graphical sensitivity analysis for the probability p that the supplier is bought and determine the critical probability values.

10.7

The multimedia company Omnismart wants to have a representative exhibition in its foyer. Companies which design inventory and booths were asked to submit and make an offer. Two companies are shortlisted, namely Kindling Design Ltd. and Bad & Fatal Ltd; Omnismart has done business with these two companies before.

As the quality of the concepts and of the construction should be approximately equal for both companies, the CEO of Omnismart, Dan Doit, wants to base his decision mainly on the two criteria of costs and time for the project. The earlier the exhibition can be opened the greater the effect on costumers and interested parties becomes.

The two companies' offers include statements about costs and duration. Doit, however, believes that strong variations from the stated expectations might occur and assigns subjective probability distributions for costs and duration.

Kindling Design Ltd.			
Duration		Costs in €	
6 weeks	15%	30,000	25%
7 weeks	35%	40,000	50%
8 weeks	35%	50,000	25%
9 weeks	15%		

Bad & Fatal Ltd			
Duration		Costs in €	
5 weeks	20%	50,000	20%
6 weeks	60%	55,000	50%
7 weeks	20%	60,000	30%

To evaluate the two offers Doit wants to use an additive value function. To test for additive utility independence, he compares the following fictitious lotteries:

1. 50% probability, 5 weeks, €60,000,
   50% probability, 9 weeks, €30,000.
2. 50% probability, 5 weeks, €30,000,
   50% probability, 9 weeks, €60,000.

He comes to the conclusion that he is indifferent between the two lotteries.

The one-dimensional utility function for the costs (x_1) which Doit defines afterwards is concave on the interval [30; 60] and can be approximated by the function

$$u(x_1) = 1 - \frac{e^{0.9} - e^{0.03x_1}}{e^{0.9} - e^{1.8}}.$$

The one-dimensional utility function for the duration (x_2) is linearly decreasing from 5 to 9 weeks. Doit defines the weighting factors to be $k_1 = 0.65$ and $k_2 = 0.35$.

(a) Which company will get the job?
(b) Perform a sensitivity analysis to determine how the decision depends on the attribute weights.

10.8

The two friends Doll and Dozy plan to improve their living standard when they retire by investing in stock funds. Each of them has €50,000 and they want to have a joint investment.

Apart from choosing the suitable stock fund they also have to decide how to split the future capital gain between them. Besides the obvious solution to split the gain evenly there is also the idea of giving the return only to one of them decided by the toss of a coin. The loser should only get his €50,000 back. The idea behind this solution is that the return (which they believe will not exceed 50%) is not large enough to provide both with a considerable improvement of their living standard.

Put yourself in Doll's (or Dozy's) position. You are certainly concerned about your own living standard but your friend's living standard is important to you as well; hence, you do not only want to consider your own gain (x_1) but also those of your partner (x_2).

(a) Test if an additive utility model is adequate by using an appropriate lottery comparison.
(b) If your answer to (a) is no: Why is this so? Can you ensure that the condition for a multiattribute utility function is fulfilled by redefining the attributes?

10.9

A community wants to build a homeless shelter and has three criteria for the selection from the different proposals: (X_1) the necessary yearly subsidy, (X_2) the number of newly created jobs and (X_3) the number of overnight stays. All three variables are uncertain.

The examination council exhibits additive utility independence of the attributes. The three one-dimensional utility functions are identified as linear over the intervals

- 100-150 thousand € (yearly subsidy)
- 8-12 (new jobs)

– 400-650 (overnight stays per month).

The first utility function decreases with rising subsidies while the two other utility functions rise with the number of jobs and overnight stays.

(a) Determine the three equations of the one-dimensional utility functions $u_r(x_r)$ (normalized for the interval [0; 1]).
(b) The council is indifferent between the sure alternative ($x_1 = 150, x_2 = 8, x_3 = 650$) and the BRL $[0.35, (x_1=100, x_2=12, x_3=650); 0.65, (x_1=150, x_2=8, x_3=400)]$.
The council also makes the following indifference statement:
$(150, 10, 400) \sim (100, 8, 400)$.
Determine the weighting factors and the utility function.

10.10

Executive Doit (Exercise 10.7.) has to place another order for a similar problem. This time, however, the criteria are not only duration of the project and costs but also the expected visitor numbers.

There are two offers *A* and *B*. *A* has a duration of 40 days and a price of €50,000, *B* has a duration of 50 days and a price of €54,000. Both bidders, however, can not guarantee these numbers. Doit thinks the realistic figures for *A* are 48 days and €55,000 and for *B* 55 days and €60,000.

When estimating the visitor numbers, Doit accounts for the fact that fewer visitors will come due to the longer construction duration but *B* also does better work which attracts more visitors. He estimates the number of visitors at 3,000 for *A* and 3,000 for *B*.

To evaluate the two offers he uses an additive function. Relevant numbers are given in the following table.

Attribute	Value range	Trend	Weight
Duration	40 to 60 days	Linear decreasing	0.2
Number of visitors	2,000 to 7,000	Linear increasing	0.5
Costs	50,000 to 70,000	Linear decreasing	0.3

(a) Evaluate the two offers.

Afterwards, Doit realizes that the estimates include significant uncertainties. He tries to depict these uncertainties in the following table.

Offer A					
Duration (days)	Probability	Number of visitors	Conditional probability	Costs (€)	Probability
40	1/3	4,500	½	50,000	1/3
		4,000	½	55,000	1/3
50	1/3	3,500	½	70,000	1/3
		3,000	½		
60	1/3	2,500	½		
		2,000	½		

Offer B					
Duration (days)	Probability	Number of visitors	Conditional probability	Costs (€)	Probability
50	1/4	5,000	½	55,000	1/4
		4,500	½	60,000	1/2
55	1/2	4,000	½	65,000	1/4
		3,500	½		
60	1/4	3,000	½		
		2,500	½		

Doit wants to transform the already defined value function into a utility function. For that purpose he imagines the following lottery: There are either 2,000 or 7,000 visitors, each with a probability of 50%. His certainty equivalent for this lottery equals 3,900.

(b) Determine the utility function. If you want to, and have a spreadsheet program, determine the optimal alternative.

10.11

A company has to choose between two investment alternatives which lead to the following net present values, depending on the state of the world also depicted in the following table. The company aims to maximize the net present value. Unfortunately, the company does not have certain information about probabilities of the states.

State	s_1	s_2	s_3
$p(s_i)$	p_1	p_2	$1-p_1-p_2$
Alternative *a*	5	8	3
Alternative *b*	6	2	6

(a) First determine analytically for which values of p_1 and p_2 alternative *a* is optimal. Illustrate your result graphically in a coordinate plane with p_1 on the x-axis and p_2 on the y-axis, and indicate the areas in which alternative *a* (respectively alternative *b*) is chosen.
(b) Now there is also an alternative *c* with the following characteristics:

State	s_1	s_2	s_3
Alternative *c*	0	4	7

How do the results from task (*b*) change because of the additional alternative?

(c) Differing from exercise (*b*), assume now that the value of alternative *c* in s_3 is 5 instead of 7. Is alternative *c* then dominated (state wise) by one of the other alternatives? Repeat your analysis from exercise (*b*) based on the modified alternative *c*. Determine the (p_1, p_2) combinations for which alternative *c* is optimal. Compare the validity of the dominance criterion with the validity of the sensitivity analysis.

10.12
It is known that an investor's utility function shows constant absolute risk aversion (i.e. it holds $u(x) = -e^{-cx}$, apart from the usual normalization). Except the fact that $c > 0$, there is no information about the investor's risk aversion coefficient *c*.

He has to choose among three projects which have the following payoffs:

- Project 1: payoff surplus uncertain, uniformly distributed for the interval [1; 2];
- Project 2: payoff surplus uncertain, two-point distribution with the realizations 1 (probability 0.4) and 2 (probability 0.6);
- Project 3: certain payoff surplus of 1.4.

Determine the optimal project choice depending on the parameter *c*. To do so, you should use an Excel spreadsheet, varying the parameter *c* in the interval [0.5; 3.5] with increments of 0.05 or 0.1. Which project and why is optimal for $c \to 0$? Which project and why is optimal for $c \to \infty$?

Case Study 1: Nine-digit zip codes

Source: Ulvila (1987), pp. 1-12.

Since 1983, the United States Postal Service (USPS) has tried to make major customers use nine-digit zip codes (ZIP+4) instead of five-digit zip codes and charged a reduced fee for this mail. At the same time, technical rationalization efforts were initiated; in particular, USPS acquired optical character recognition devices (OCRs) and barcode scanners (BCSs). OCRs read the zip code and print the corresponding barcode on the mail. BCSs then read the barcode and sort the letters. Because the nine-digit zip code allows for precise addressing, manual sorting is no longer necessary. In 1984, the first stage of these measures was completed. For the second stage, the acquisition of a further 403 OCRs and 452 BCSs was planned with a total contract volume of $450 million. At that point, the US Congress asked the Office of Technology Assessment (OTA) to verify whether the continuation of the program was advisable. The OTA brought in a consulting firm for the analysis.

Over the course of the analysis, it became apparent that the choice was not simply a straight one between the implementation or omission of the second stage; more alternatives were worth looking at. The OCRs that had been purchased so far and those that were going to be purchased were single-line devices capable of reading the last line of the address (which is usually the line that contains the zip code). However, other devices were available that were capable of reading up to four lines. For mail with five-digit zip codes, these devices had the advantage that they could often read the address on the envelope, use this information to extract the right nine-digit zip code from a database and print the corresponding barcode on the envelope.

Altogether, six *options* were identified:

A. The acquisition of further OCRs, as planned, in particular only single-line OCRs,
B. The acquisition of multi-line OCRs and conversion of current devices into multi-line devices (while retaining nine-digit zip codes),
C. The conversion to multi-line readers as in option *B*, but abolition of nine-digit zip codes,
D. The acquisition of planned devices as in option *A*, but research and testing of ways of conversion and the replacement of all devices with multi-line OCRs as soon as possible,
E. As in option *D*, but in two steps. The conversion to multi-line activity would only take place if there was little or no use of nine-digit zip codes until the end of 1987, or,
F. No acquisition of further devices and the abolition of nine-digit zip codes.

The only *objective* of the decision was to reduce costs. The principal objective variable that was considered was the NPV of the cash flow series over the planning horizon (1985–1998), discounted at an interest rate of 15%. The main *uncertainties* concerned (1) the extent to which the nine-digit zip codes were used, (2) the savings potential of the devices and (3) the actual savings of the multi-line devices, depending on their utilization. For each of these continuous variables, a discrete distribution was estimated. The best and worst values were allocated a probability of 18.5% and the middle value received a probability of 63%. Consequently, a decision tree was obtained with nine consequences for each of the options *A* and *C* and 27 consequences for each of the options *B*, *D*, and *E*. Option *F* was used as a standard of comparison for the other options.

For each consequence and each year of the planning horizon, cash in- and outflows were estimated on the basis of a detailed analysis. The cash flows were then discounted to determine the NPVs. The calculations were computed using a common spreadsheet software package. The following table shows an example of the calculations for the consequence for option *D* when all three uncertain variables are at their respective best possible values. The numbers are stated in $1,000.

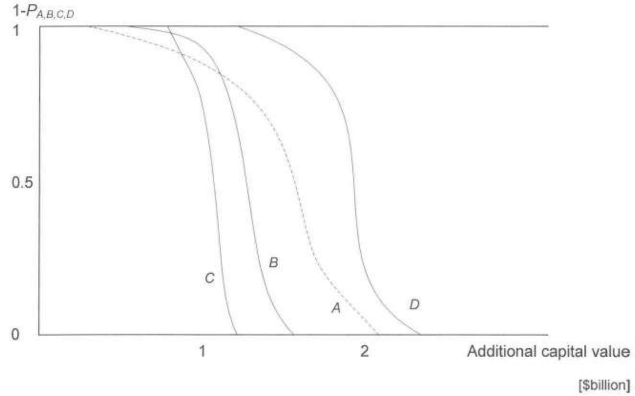
Jahr	1985	1986	1987	1988	1989	1990	1991	...	1998
One-line OCRs	–140,325	–140,325	–113,200						
Add. costs	–20,077	–20,077	–16,147						
Research expenses	–5,000	–5,000	–5,000						
Conversion equipment				–43,667	–43,667	–43,667			
Maintenance and spare parts	–27,706	–34,172	–36,640	–42,950	–57,685	–98,690	–86,221		–142,301
Address information	–32,400	–30,900	–15,700	–16,700	–18,000	–19,336	–20,770		–34,280
Investment and maintenance - total	–225,508	–230,474	–186,687	–103,317	–119,352	–161,693	–106,991		–176,581
Reduction of fees	–58,333	–88,667	–117,444	–136,111	–140,000	–140,000	–140,000		–140,000
Saving of wages	161,128	268,953	465,747	711,526	827,240	888,621	1,015,649		1,676,253
Cash surplus / Pos. net cash flow	–122,713	–50,188	161,616	472,098	567,888	586,929	768,657		1,359,671
Additional surplus compared with *F*	–177,182	–85,072	128,190	436,381	536,762	553,493	732,741		1,300,394

This consequence has a surplus NPV of $2.6 billion or $2.4 billion when compared with option *F*. Using the probability distributions of the uncertain variables, the following *expected values* of NPV were obtained:

A	$1.3bn
B	$1.2bn
C	$0.9bn
D	$1.5bn
E	$1.4bn

D had the highest expected value; this was the conversion of one-line devices into multi-line devices. All options were better than option *F*, which amounted to dropping the program.

To analyze the *risk*, the distributions of the NPVs were compared with one another. It became clear that option *D* was absolutely dominant toward option *E*, i.e. for every possible data set-up, option *D* had a higher NPV. Hence, option *E*, which was the two-step alternative, could be ruled out.



The cumulative probabilities for the NPVs were then identified. In a smoothed form, the results could be depicted as *risk profiles* (see figure above). Fortunately, one alternative stochastically dominated the remaining alternatives: for option D, the probability of exceeding any randomly chosen NPV (within the limits) was higher than for any other option. Hence, option D was recommended without needing to know about the risk attitude of the decision maker. This analysis was the basis for the report of the OTA to the US Congress in June 1984.

Case Study 2: Stockpiling of a blood bank

Source: Keeney and Raiffa (1976), pp. 275–281.

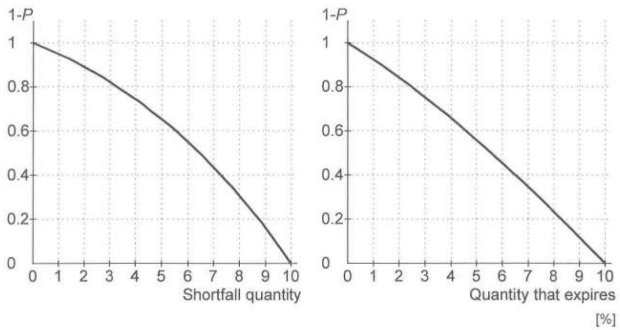
The stockpiling of stocks of individual blood types is complicated because the demand for them cannot be predicted with certainty. A low stock can lead to shortfall quantities, which can in turn have very unpleasant consequences (rush orders, phoning up professional blood donors, postponement of surgeries). Overstocking, however, can lead to stored blood becoming useless because of a violation of the expiration date.

In order to rationally stockpile, it is necessary (among other things) to determine the utility function of the two attributes "shortfall quantity" *(S)* and "quantity that expires" *(E)*. In a study at the Cambridge Hospital (Cambridge, MA, USA), the utility function of the nurse responsible for blood orders was determined. The range of values of S was set between 0 and 10% of the demand. The range of E was set between 0 and 10% of the quantity stored over the whole year.

After the validity of the condition of mutual utility independence had been ensured, the two utility functions $u_S(s)$ and $u_E(e)$ were determined in the following way. First, some of the points of each utility were determined, which were then approximated by the functions

$$u_S = 1 + 0.375\left(1 - e^{0.13s}\right)$$
$$u_E = 1 + 2.033\left(1 - e^{0.04e}\right).$$

These functions are depicted in the figure below. As you can see from the figure, the decision maker is risk averse with respect to both attributes.



To determine the weight factors k_S and k_E the following values $u(s,e)$ were defined:

$$u(0,0) = 1$$
$$u(10,10) = 0$$
$$u(10,0) = k_S$$

$$u(0,10) = k_E.$$

According to the trade off method, the nurse decided that she was indifferent between (0, 10) and (4.75, 0). It must hence hold that

$$u(4.75,0) = u(0,10) = k_S.$$

Written in the multiplicative form:

$$u(4.75,0) = k_S \cdot u_S(4.75) + k_E \cdot u_E(0) + (1 - k_S - k_E) \cdot u_S(4.75) \cdot u_E(0) = k_S$$

If you insert $u_E(4.75) = 0.68$ and $u_S(0) = 1$, you obtain

$$k_S = 0.68 + 0.32 \cdot k_E.$$

A second equation was now needed. This equation was obtained by the lottery method. Using the nurse's indifference statement

$$(6,6) \sim [0.5,(10,10);0.5,(0,0)]$$

it follows that $u(6, 6) = 0.5$. It thus holds that

$$k_S \cdot u_S(6) + k_E \cdot u_E(6) + (1 - k_S - k_E) \cdot u_S(6) \cdot u_E(6) = 0.5.$$

Inserting $u_S(6) = 0.56$ and $u_E(6) = 0.45$ leads to

$$0.308 \cdot k_S + 0.198 \cdot k_E = 0.248.$$

From these two equations, it can be seen that $k_S = 0.72$ and $k_E = 0.13$. The two-dimensional utility function is hence

$$u(s,e) = 0.72 \cdot u_S(s) + 0.13 \cdot u_E(e) + 0.15 \cdot u_S(s) \cdot u_E(e)$$

This function can then be used to evaluate all uncertain consequences. For a particular order policy, we assume that three different consequences are possible. These consequences and their respective probabilities are described in the following table. This table also contains the utilities of the consequences and determination of the expected utility value of this order policy (0.624).

Consequence i	Probability p_i	Shortfall quantity s_i	Quantity that expires e_i	Utility $u(s_i,e_i)$	$u_i p_i$
1	0.25	8	2	0.373	0.093
2	0.60	3	6	0.705	0.423
3	0.15	0	10	0.720	0.108
Total	1.00				0.624

Chapter 11: Time preferences under certain expectations

11.0 Summary

1. In many decision-making processes you will have to weigh immediate against future consequences. In these processes, the time reference of the consequences must be taken into consideration.
2. Intertemporal decisions can be understood as multiple objective decisions with additive value functions. The i-th attribute is replaced by the i-th point in time.
3. The prerequisite of mutual preference independence for an additive intertemporal value function is not always given. If this prerequisite cannot be met, redefining the system of objectives generally helps.
4. The intertemporal value function can be simplified to the discounting or the Harvey models if - apart from the condition of mutual preference independence - other axioms are fulfilled as well.
5. For the projection of temporally inconsistent and possibly undesirable decisions you can use the quasi-hyperbolic discounting model and take preventive measures accordingly.
6. Whether you wish to use one of these simple models depends on your acceptance of the proposed axioms. Should you find all of the axiom systems inappropriate for your process of rational decision-making, you should take recourse to the general form of the (additive) intertemporal value function.
7. Should you have options for temporal transformation, as, for instance given in a perfect capital market, you can use the discounting model independent of axiomatic considerations and personal preferences.

11.1 The problem of time preference

Frequently you will find yourself having to make decisions in which you have to weigh immediate sacrifices or benefits against future sacrifices or benefits. These decisions may relate to monetary or non-monetary factors. Some examples will illustrate the range of decision situations in which temporal factors play a part:

You wish to change your diet to whole foods because you know this will make you live more healthily, feel better and enjoy life more. These benefits will not be immediately available but will take effect over a timeframe of weeks, months, or years. The inconveniences of a change in diet will however be felt instantly: being used to a normal diet, you do not like wholefood meals. You have to make yourself stick to your new healthy diet plan. You have to suppress your appetite for meat and sugar. Food preparation will take longer and shopping for food may involve more traveling. You are considering whether you want to put up with all this.

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_11, © Springer-Verlag Berlin Heidelberg 2010

The government could allocate financial resources for research into renewable energies in order to make clean energy resources sustainable for the future. As a politician, you are involved in the decision-making process. Again, the immediate drawbacks are evident: there will be less money in your limited budget for other pressing tasks (new preschool places, purchase of tanks, staff cars etc.). The benefits will only take effect at a later date when the renewable energies will be cost-efficiently usable and the plants (wind turbines, solar arrays) have been built and put into operation.

These decision situations have one thing in common: the consequences that have to be taken into account take effect at different points in time; in this case, we speak of *intertemporal* decisions. In this chapter, we will have a look at how to behave rationally in the case of intertemporal decisions with a narrow focus on intertemporal decisions under certain expectations. It would of course be more realistic to also consider models for intertemporal decisions under risk. The complexity of the pertinent models and the unsatisfactory state of research, however, have led us to omit these approaches. Frederick et al. (2002) present an excellent overview of prescriptive and descriptive research on intertemporal decisions.

11.2 The additive intertemporal value function

11.2.1 Derivation of the intertemporal value function

If the consequences that need to be taken into account take effect at different moments, intertemporal decision-making can be understood as a problem of multiple objectives. Imagine you are a manager in charge of a profit center. Your main objective is profit. This main objective can be subdivided into the secondary objectives of profit in 2010, 2011, 2012 etc.

By devising such a system of objectives, you formally equate multiple objective decisions to intertemporal decisions; this hence tells you how to solve the intertemporal decision problem: using the methods you know from Chapter 6, you can now make a rational decision. In the case of intertemporal decisions, the multiattribute value function is called intertemporal value function.

Let $(a_0, ..., a_T)$ be the vector of intertemporal consequences of an alternative a. The point in time at which the consequences will occur are numbered from 0 to T; 0 refers to the present, T designates the end of the planning period, i.e. the point in time furthest in the future. Chapter 6 showed under which conditions the multiattribute value function v can be represented by a simple additive form. If the preference of the decision maker is transitive and complete and meets the condition of mutual preference independence (MPI), the preference can be represented by the additive intertemporal value function:

$$v(a) = \sum_{t=0}^{T} w_t v_t(a_t). \tag{11.1}$$

The intertemporal value function can be measured if the preference also complies to the condition of difference independence. In the following, we will assume that this condition is met, i.e. we will only look at measurable intertemporal value functions.

Just as in the case of a multiattribute value function there is one single value function for each attribute, the decision maker can attribute one single value function v_t to each point in time t. The period value functions can be normalized to attribute the value 0 to the worst and the value 1 to the best possible outcome for any particular period. To determine the period value function, apply the methods that you know from Chapter 5. Later in this chapter we will from time to time argue that an alternative may have "no consequence" at a specific point in time. This, of course, is not permitted if you interpret an intertemporal problem as a multiple objective problem. In these instances, our formulation should be interpreted to refer to a consequence that does not offer any extra benefits (and for reasons of simplicity we will assume that this generally applies to consequence 0, the consequence which does not alter the status quo). The weights w_t will be determined by using one of the methods introduced in Section 6.4. The weights w_t describe the trade-offs between individual periods, i.e. they represent the importance the decision maker assigns to each unit of the corresponding period value function. For intertemporal decisions, the value w_0 (weight of period 0) is routinely set to 1. In return, the requirement that the period weights have to add up to 1 is abandoned. You can normalize $w_0 = 1$ since value functions can be transformed linearly.

Let us once again explain this procedure which in principle should be clear from the preceding chapters, using the example of a profit center manager. His target values are the profits (in million €) within the next three years. The manager compares the profit vectors

$$a = (50, 50, 100) \quad \text{and} \quad b = (30, 70, 100).$$

In both sets, the profits are the same in the third year. The MPI requirement demands that the manager's preference must not change between sets if the profit in year three is changed in the same manner for both alternatives. If the manager for instance prefers $a \succeq b$ then accordingly,

$$a' = (50, 50, 200) \succeq b' = (30, 70, 200)$$

and

$$a'' = (50, 50, 50) \succeq b'' = (30, 70, 50)$$

must apply. Bear in mind that the MPI requirement does not prescribe any preference between a or b; the MPI requirement simply says that this preference cannot depend on periods in which both alternatives have the same consequence. The MPI requirement seems plausible but there are many situations in which it simply will not be applicable. We will have a closer look at this problem in Section 11.2.2.

If mutual preference independence applies, the period value functions $v_t(x_t)$ for the profits x_t can be determined. Our manager can use the methods known from Section 5.2: the direct-rating-method, the difference standard sequence technique, or the bisection method.

The ranges of definition for individual period value functions may differ: if for period 2 the highest possible profit is 170 and the lowest possible profit is 30, v_2 must be determined for the corresponding range. This shows that individual period value functions may also differ. In their individual range of definition they can always be normalized to lie between zero and one. Should the worst possible consequence occur, it will not contribute to the value of the whole set: $v_t(x_t^-) = 0$. The weighting can be carried out by the same methods that are used for multiple objectives. To illustrate this: using the swing method, one would assume a fictitious alternative which would represent the worst possible outcome in each period. The transitions to the best outcomes in each period would then be sorted and evaluated on a cardinal scale.

The manager's aims (profits of the centre within the next three years) are purely monetary and differ only with regard to their respective points in time; this does not always have to be the case. If you consider changing your diet to whole foods and - to simplify matters – allow only for two points in time ("today," and "in future"), you may identify today's goal as "delicious food" and your future goal as "health." These aims are not monetary, refer to different moments in time, and to different contents. The same holds for the decision of investing in renewable energies or new preschool places: depending on the definition of the problem, the objectives for each period may change. In a lot of instances, more than one attribute will be relevant per period. Consider, for instance, the choice of your first employment after graduation. In Chapter 5 and Chapter 6 you were introduced to the single-period version of the problem for one or multiple objectives (salary and working hours). More realistically, however, the consequences of each employment should be projected down to the end of your planning horizon. For each point in time of your planning period an alternative is described by its outcomes for all relevant attributes. In order to obtain the intertemporal value function in this case, you have to evaluate the consequences for each point of time with the multiattribute value function $v_t(x_{1t}, x_{2t}, \ldots, x_{mt})$.

Defining the consequences one should consider that a consequence has a value not only at the time of its occurrence but may also have a value in the past and in the future (Berns et al. 2007, p. 483 ff.). A future consequence may today provoke fear or anticipation. If, for instance, you are toying with the idea of buying a yacht two years from now, the mere thought of owning a yacht will probably make you happy today. This effect will not be represented by the attributes assigned to the period in which you will own the yacht. Consequently, you need to define a new objective variable, something like "looking forward to future sailing trips." In the same manner, consequences may affect the future: somebody who has sailed around the world will probably draw on the experience for the rest of his life. The lasting effects of joy or sorrow can be described by the extra objective "life satisfaction up to today".

11.2.2 Discussion of assumptions for the additive intertemporal value functions

The condition of mutual preference independence is a prerequisite for an additive model. As shown in Chapter 6, this condition is not always given. For intertemporal decisions, this assumption for the additive model needs to be discussed again.

Suppose that a manager has the following preference with regard to his profits:

$$(50, 50, 100) \prec (30, 70, 100).$$

He prefers the second alternative to the first one because the second one includes an upward trend. He also has a second preference:

$$(50, 50, 20) \succ (30, 70, 20).$$

In this comparison, he prefers the first alternative because here the decrease in year three is not as dramatic as in the second one. The manager's preference however violates the condition of MPI as the profits for year one are identical in both pair comparisons (Eisenführ 1988). The set of profits within the period affects the preference: the profits of one single period cannot be valued independently of the profits in the other periods, i.e. the MPI requirement is not fulfilled. The dependence of a preference on a trend that is reflected in the consequences has been corroborated by empirical studies (for an overview, see Read and Powell 2002).

Loewenstein (1987) conducted a study in which the condition of preference independence was tested directly. The study presented students with four alternatives suggesting different ways of spending the next three Saturday nights.

Table 11-1: Four alternatives for the next three Saturday nights

Alternative	This weekend	Next weekend	Weekend after next	Preference
a	eat out, French	eat at home	eat at home	16%
b	eat at home	eat out, French	eat at home	84%
c	eat out, French	eat at home	have lobster	57%
d	eat at home	eat out, French	have lobster	43%

Given the choice between *a* and *b*, a significant majority opted for alternative *b* (84% against 16%), given the choice between *c* and *d*, however, a slight majority preferred *c* (57% against 43%). Note here that *a* and *b*, just like *c* and *d*, are identical for the weekend after next. Consequently, under the assumption of MPI, the weekend after next must be irrelevant for the decision. Since *a* and *c*, just like *b* and *d*, are identical on the remaining relevant weekends, a decision maker must prefer *a* and *c*, or *b* and *d* (indifferent statements were not permitted). From the results of this study we must conclude that preference independence is not always given.

If the MPI requirement is not fulfilled, you can try to change to another system of objectives. In the example given earlier, such a system might include the average annual profits, the profit growth rate, and the distribution of the profits; the new system might then comply with the MPI rule.

Should it prove impossible to find such an optimized system of objectives, one could alternatively choose an approach in which the outcomes that have been evaluated using period value functions are linked non-additively. Loewenstein and Prelec (1993), for instance, propose such a non-additive aggregation, which takes into account not only the values of a sequence but also their trends and the uniformity of that trend. In this model, anticipation of future events has an impact on the preferences of previous periods. Apart from the additional mathematical effort required here, the model also rests on questionable normative assumptions. Particularly in reply to the question whether increasing, decreasing or unchanging outcome sequences are preferable, a number of good reasons can be found to support each pattern.

Anticipation and contrast effects speak in favor of increasing sequences if the results are compared to previous periods and the increasing sequence thus leads to continuous improvement.

Decreasing sequences could be preferred if uncertainty with regard to the actual occurrence of future events is assumed and it is uncertain whether the best result will actually be realized in a later period. Likewise, delay of payments could cause opportunity costs since the money could have been invested profitably in the meantime.

Unchanging sequences may be attractive if consumption shows diminishing marginal utility, i.e. if an additional unit of consumption creates little added benefit when a high basic level has already been achieved.

Even from a descriptive point of view, there is no conclusive answer to the question which form of trend will be preferred. Frederick and Loewenstein (2008) have shown that simply changing the question can produce preferences for increasing, decreasing and unchanging sequences. It therefore seems advisable for rational decisions to have a closer look at the reasons for the attractiveness of such patterns and to adapt the system of objectives in such a way that the important motives for a specific situation are accurately represented.

Overall, we must conclude that in many intertemporal decision processes, the assumption of preference independence is problematic, making the use of the additive model questionable. On the other hand, alternative models rest on dubious assumptions as well. For our further considerations, let us assume that we have managed to find a system of objectives that complies with the MPI requirement.

11.3 Special types of the additive intertemporal value function

11.3.1 Identical value functions for each period

The additive intertemporal value function can be simplified if the preference fulfills further axioms apart from the MPI requirement. We will only present a few reflections on the axiomatic principles of time preference here, more detail can be found for instance in Ahlbrecht and Weber (1995). Considering the additive intertemporal value function (Formula (11.1)), it becomes clear that the value functions of individual periods can differ significantly. Let us therefore first have a look at the conditions that need to be fulfilled to imply identical value functions for each period. For many applications, this is a desirable and significant simplification; just think of the manager who has to assess the profits of his center.

It would surely be easiest to simply assume identical value functions. In keeping with the axiomatic approach of this textbook, however, you will learn one condition which allows us to derive identical value functions.

Definition 11.1 (constant preference differences)
Let a, b, c, and d be alternatives, and s and t any two points of time, then *constant preference differences* are given if

$$(a_s \to b_s) \succeq (c_s \to d_s) \quad \Leftrightarrow \quad (a_t \to b_t) \succeq (c_t \to d_t),$$

where x_t denotes the intertemporal consequences that will occur for alternative x at point t.

The condition of constant preference differences states that the ordering of differences in preferences between alternatives is independent of the relevant point of time. Should you, for instance, favor the transition from €1M to €2M in next year's profits over the transition from €14M to €15.5M, then this preference must be valid if you compare these profits at a later date as well.

We can now show that the condition of constant preference differences is equal to the existence of identical value functions, i.e. the intertemporal value function assumes the form (11.2) if the condition for constant preference differences holds.

$$v(a) = \sum_{t=0}^{T} w_t v_0(a_t). \tag{11.2}$$

We denote the period value function that is identical for all periods by v_0, and thereby express that the function is easiest to determine for the point $t = 0$. The weights w_t which have so far not been specified can be expressed in a manner that demonstrates their connection with the discounting model as a special form of the intertemporal utility function. We write:

$$w_t = \frac{1}{(1+i)^{\alpha(t)}}. \tag{11.3}$$

Time preference is described by a constant discounting rate $(1 + i)$ and the factor $\alpha(t)$. Without the additional transformation of time t by the function α, i.e. for the case $\alpha(t) = t$, the result would be the classic discounting model in which the period weight falls exponentially (at the rate of $1 + i$) in time. The more general representation by $\alpha(t)$ allows us to posit that the level of discounting does not depend on the actual time t in which the consequence occurs but on how far into the future time t is *perceived*. A decision maker will probably perceive periods 101 and 102

to be closer together than periods 1 and 2. This can be illustrated by ensuring that $\alpha(102) - \alpha(101)$ is less than $\alpha(2) - \alpha(1)$.

Various specific models may now be represented by different forms of the function $\alpha(t)$. In the following, three of these forms will be introduced. For all of them, we will assume that the points of time in the intertemporal alternative are equidistant, i.e. that they represent the endpoints of consecutive periods of equal length.

11.3.2 The discounting model

An additive intertemporal value function describes a preference that is complete and intransitive and which fulfills the condition of mutual preference independence. If the preference also fulfills the stationarity axiom (which will be explained in a moment) the intertemporal value function will take the following form (Dyckhoff 1988, p. 1002, French 1988, p. 135):

$$v(a) = \sum_{t=0}^{T} \frac{1}{q^t} v_0(a_t). \tag{11.4}$$

The period function that is identical for all periods has here been denoted by v_0. In the discounting model, the period weights w_t of the additive intertemporal value function are simplified to powers of a discounting factor $q = (1 + i) > 0$. As indicated before, $\alpha(t) = t$ remains valid in this case, i.e. time is perceived to be linear. The discounting factor can thus be determined easily. If a decision maker is indifferent between a consequence a_0 due in period 0 and a consequence a_1 due in period 1, the discounting model will yield $v(a_0, 0) = v(0, a_1)$, or

$$v_0(a_0) = \frac{v_0(a_1)}{q}, \text{ i.e. } q = \frac{v_0(a_1)}{v_0(a_0)}. \tag{11.5}$$

If the period value function is known, the discounting factor q can be calculated directly. As in all methods of determination, systematic biases and errors may occur when determining the discounting factor. It is therefore advisable to have the decision maker make several indifference statements. Possible inconsistencies can thus be eliminated by repeated interviews or averaging, as demonstrated in Chapter 6.

In order to be able to formally define the stationarity axiom, we have to introduce the notion of shifting a sequence of outcomes.

Definition 11.2 (shifting a sequence of outcomes)
Shifting a sequence of outcomes by n periods ($n \geq 0$) relocates each consequence a_t of alternative a to period $t + n$. Shifting alternative a by n periods will be called δ_n and the new alternative is accordingly denoted by $\delta_n(a)$. Figure 11-1 shows the shifting of an alternative by two periods.

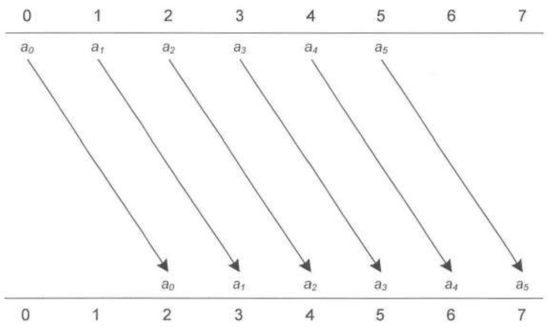


Figure 11-1: Shifting a sequence of outcomes by two periods

Definition 11.3 (stationarity)
With regard to sequences of outcomes, a preference is called stationary if for all sequences a, b, and any shifting length n, preference $a \preceq b$ holds if and only if after the shifting δ_n preference $\delta_n(a) \preceq \delta_n(b)$ holds. Stationarity means that the preference is preserved after shifting.

A simple example illustrates the stationarity axiom. You are planning next year's vacation. Your job responsibilities will allow for only one holiday per year, so you are wondering whether to go on vacation during summer or next winter. Suppose you love skiing and opt for winter. Alternatively, suppose you are planning your vacation for the year after next and you are not going away this year. The condition of stationarity now requires that your decision for the year after next will be identical to the one for next year, i.e. that you favor the winter vacation in both cases.

The condition of stationarity allows us to simplify the additive intertemporal value function to the discounting model. However, stationarity cannot be regarded as a general prerequisite for rational behavior. Why not prefer the winter vacation for next year but warmer climates for retirement? In the discounting model, period value functions are identical for all periods and period weights have a very special form. This narrows down the range of possible preferences. At this point, the argument must be made similar to that for decisions under risk. For the latter, the expected value of an objective variable is easier to calculate, but as it narrows down the range of possible preferences so significantly, the more general expected utility theory should be accepted as a basis for rational decisions. In some intertemporal decision situations it will be appropriate to use the discounting model while in other situations the more general additive temporal model must be used. It is therefore essential for you to become more familiar with the condition of stationarity so that you are able to choose the appropriate model for each decision problem.

The condition of stationarity requires that preferences of alternatives do not change when the consequences of these alternatives are shifted simultaneously. Decision makers that do not evaluate future consequences according to the discounting model may therefore behave inconsistently and retrospectively reverse previous decisions when reviewing the problem using the same model (Strotz 1957).

Under real conditions, shifting a sequence of outcomes will usually require a complete remodeling of the problem. In general, decisions can only be understood within the temporal framework in relation to which they are made. Months later, the same problem will present itself in a different light, for instance because there are new alternatives to be considered or because new interests have developed which require a new system of objectives. Maybe you have been shaken up by a roadshow of your local forest society about the negative effects skiing has on the animal and plant life in the mountains. "Environmentally safe vacation" would then become a new goal in your system of objectives. A change of mind in favor of the summer vacation would be quite possible. Not only do the consequences of either summer or winter vacation seem to be displaced, but new consequences have also been added. Decisions that have to be made at different times will usually vary significantly.

Another term that is important for our further considerations is the notion of impatience. Impatience can be formally defined: let a and b be two alternatives that both have consequences only today (in period zero). The decision maker expresses his preference $a \succeq b$, i.e. he likes the consequence of a in period zero at least as well as the consequence of b in period zero. Additionally, s and t will denote any two points in time, with t greater than s. Based on a and b, two new alternatives a' and b' can be constructed that will only have consequences at the two points of time s and t. Table 11-2 shows the alternatives a' (the better or equally good consequence occurs first) and b' (the worse or equally good consequence occurs first). A decision maker is considered impatient if in this situation he always prefers a' to b'. Impatience implies that the factor q in the discounting model is larger than one.

Table 11-2: For the definition of „impatience“

Alternative	Consequence in s	Consequence in t
a'	a	b
b'	b	a

Under the assumption of impatience, a time dominance criterion can be defined analogously to the principle of stochastic dominance in decisions under risk. An alternative a dominates a second alternative b if the sum across all previous, evaluated consequences of a is at least as large at all points in time as that of b and is truly larger at least at one point in time. If the evaluated alternatives are sufficiently different or the dominance criterion is selective enough, the best alternative can be identified (Mus 1988). For a comparison of different dominance criteria, see Dyckhoff (1988, p. 991 ff.).

Impatience cannot necessarily be regarded as a basis for rational behavior. A priori, there is no compelling reason to assume that immediate consequences are more important than future ones. Why should a gummy bear now (or certain immediate profits of a profit center) necessarily be preferred to tomorrow's gummy bear (or identical profits of the center next year)? (Should you think a gummy bear is too small to be even considered, imagine a bagful of gummy bears. Should you belong to the minority that does not like gummy bears, think of a night at the opera.) To accept the assumption of impatience ultimately remains - as for every decision maker - up to you.

11.3.3 The Harvey model

In addition to the discounting model, there are numerous approaches trying to model time preferences. Some of those are full models, some are just extensions of existing models. The hyperbolic discounting model can be viewed as the most prominent of these approaches.[1] We will come back to the (simpler) case of quasi-hyperbolic discounting in more detail in Section 11.3.4. Most of the models are not trying to provide prescriptive advice for decision making but instead trying to describe actual decision making and thus model systematic deviations from rational behavior. An important concept in this context is time inconsistency, i.e. differing preferences over time, which lead to a reversal of decisions made in the past once time proceeds.

You are possibly asking yourselves why we are discussing these kinds of models at all if our primary interest lies in prescriptive decision theory. The reason is that, while making rational decisions, it may be useful to be aware of the formulation of our preferences. Being aware of temporal inconsistencies can help take measures that facilitate desirable long-term behavior and impede undesirable behavior. Take for example, your shopping at your local supermarket in preparation of a relaxed evening in front of the TV: either you buy as much candy as you wish to eat that night or you go right ahead and buy a little more to keep for the next night. While alternative 2 is probably preferable for time consistent preferences (lower costs because of family packs, no extra trip to the supermarket the next day), your knowledge of quasi-hyperbolic discounting (in other words, knowing about your lack of discipline) would probably tell you to decide otherwise. If you were to buy the larger amount of sweets and in the evening were faced with the decision to either only consume the planned amount or to consume a bit more, the high preference for immediate consumption will tend to result in overconsumption since the future disadvantages carry only a comparatively light weight. Knowing this, you now can in the supermarket already make the (from this superior perspective, rational) decision to choose only small packs in order to discipline yourself in the evening.

[1] Compare in detail Angeletos et al. (2001).

The Harvey model can be seen as a bridge between prescriptive and descriptive models. It yields quite convincing results with regards to describing actual dynamic behavior. Additionally, it is well suited to demonstrate how decision models can be derived from axioms that help the decision maker evaluate his own acceptance of a certain model. Harvey (1986) was able to demonstrate that, if the decision maker's preference fulfills not only the MPI condition but also the condition of impatience and the – still to be defined – axiom of invariance under stretching, the intertemporal value function has the form shown in Formula (11.6).

Here, the period value functions are also identical for all periods. We adopt the value function of period zero v_0 for all further periods. The period weights are the inverse of the period number plus one to the power of r ($r \geq 0$):

$$v(x) = \sum_{t=0}^{T} \frac{1}{(1+t)^r} v_0(x_t). \tag{11.6}$$

The function $\alpha(t)$ in the Harvey model therefore reads:

$$\alpha(t) = r \cdot \frac{\ln(1+t)}{\ln(1+i)}. \tag{11.7}$$

This function $\alpha(t)$ clearly shows that time here (unlike in the discounting model) is not linear but logarithmical. The concavity of the function $\alpha(t)$ implies that a time interval is perceived to be shorter the further it is shifted into the future.

Because of this difference in time perception, there is only one specific case in which the Harvey and the discounting model coincide: when no discounting takes place at all. Here, the total evaluation of a sequence of outcomes is simply the sum of the period evaluations (this is true for the parameters $q = 1$ in the discounting model and $r = 0$ in the Harvey model, making logarithmical time perception irrelevant).

In all other instances, the coefficient r in the Harvey model becomes a linear factor in the time perception. Like the parameter q in the discounting model, it must be established by interviews. For this purpose, any consequence x_0 can be specified. We will now first determine x'_0, which makes the decision maker indifferent toward the transitions from value zero to x_0 and from x_0 to x'_0. It holds that $2 \cdot v_0(x_0) = v_0(x'_0)$. The decision maker will then be asked for the period t' in which he will be indifferent between the immediate consequence x_0 and the consequence x'_0 in period t'. From this indifference statement follows that period t' has half the weight of period zero. It is therefore also referred to as the half-value period. The half-value period will directly yield the parameter r:

$$r = \frac{\log 2}{\log(1+t')}. \tag{11.8}$$

When determining the parameter r, we again need to check for consistency. The easiest way is to ask the decision maker for the period t'' in which consequence x_0 has only a third of the value it has in period zero. If, in Formula (11.8), the number 2 in the numerator is replaced by the number 3, and in the denominator t' is replaced by t'', consistent decision makers should obtain the same parameter r. Potential inconsistencies can be resolved by feedback or averaging.

Let us now define the axiom of invariance under stretching, on which the Harvey model is based.

Definition 11.4 (stretching a sequence of outcomes)
We call β_n a stretching by factor n and $\beta_n(a)$ the result of stretching alternative a. The stretching β_n of a sequence of outcomes a shifts consequence a_i of alternative a to period $(i + 1) \cdot n - 1$.

Figure 11-2 shows stretching by the factor 2. The axiom of invariance under stretching states that for all (intertemporal) alternatives a and b and any stretching β_n preference $a \prec b$ holds if and only if $\beta_n(a) \prec \beta_n(b)$ also holds.

We will illustrate the axiom of invariance under stretching with the following example. Suppose you are indifferent between a beach vacation in three months' time and a city trip in six months from now. In order to fulfill the axiom, you must still be indifferent between the beach vacation in ((3 + 1) · 2 – 1 =) seven months time and the city trip in ((6 + 1) · 2 – 1 =) thirteen months time (stretching by factor 2); similarly if the waiting times are 11 and 20 months (factor 3).

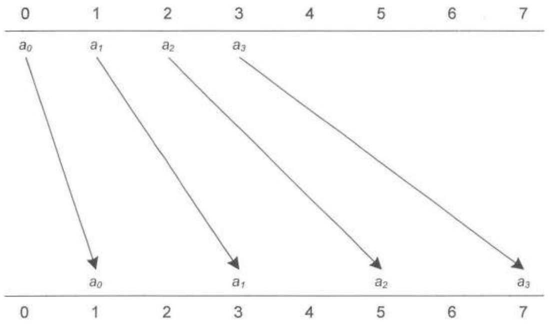


Figure 11-2: Stretching of a sequence of outcomes by the factor 2

As we can see, the condition of invariance under stretching is not a necessary condition for rational behavior. If you however accept this condition and the axiom of impatience, the resulting intertemporal value function is very simple.

11.3.4 A comparison of the two axiom systems and an alternative model

In 11.3.2 and 11.3.3. we have so far introduced two special cases of the additive intertemporal value function; both are based on similar axioms. The Harvey model

assumes impatience and invariance under stretching whereas the discounting model requires stationarity, which can also be called invariance under shifting.

Which of the two axiom systems is preferable? First of all, remember that as a decision maker, you can clearly make rational decisions on the basis of the general additive intertemporal value function – you do not necessarily have to turn to either of the two special cases. Frederick et al. (2002) introduce further models as well as extensions (which in part can be derived from the general form introduced in Section 11.3.1), and a model with different period value functions can of course also be employed. The main advantage of both the discounting and the Harvey model is their simple form: the period value functions are identical, and in both models only one weighting parameter has to be identified.

Although it requires two parameters to identify the weighting function, the model of quasi-hyperbolic discounting, first introduced by Phelps and Pollak (1968) and improved by Laibson (1997), also enjoys the advantage of simplicity. Axiomatization would also be quite easy as the condition of stationarity would only need to be changed slightly. The model is however more descriptive and rests upon the observation that frequently, neither the axiom of stationarity nor the axiom of stretching are fulfilled in their pure forms. By contrast, there is generally a strong preference for immediate outcomes in intertemporal decisions, i.e. immediate consumption will receive the value of 1 while any consumption anticipated in the future will enter the utility calculus with a substantial discount. There are of course also differences between the weights for future periods, but the most significant discounting is between periods 0 and 1. As a result, this reflects a high degree of impatience. Formally, the quasi-hyperbolic discounting function can be written as:

$$v(a)=v_0(a_0)+\beta\sum_{t=1}^{T}\delta^t\cdot v_0(a_t) \tag{11.9}$$

The parameter β represents the preference for immediate outcomes and has a value between 0 and 1 while δ expresses the common discounting for periods in the future. Note that β only influences the weighting for later periods (starting with $t = 1$), thus "devaluating" future events independent of their distance in the future. For the special case $\beta = 1$ the discounting model applies with $q = 1/\delta$.

Similar to the Harvey model, the quasi-hyperbolic discounting model is can represent situations in which individual behavior is time inconsistent. A typical example to which it can be applied is the problem of many smokers. They desire, for the sake of their health, to stop smoking – and a great many regularly plan to stop in the future (i.e. to do without future pleasures of smoking to enjoy health benefits even further in the future). At the moment the cigarette beckons, however, the utility of the immediate enjoyment of smoking will enter the utility function completely (i.e., without β-devaluation) while the future health benefits will be substantially discounted. The smoker decides in favor of immediate consumption. At the same time, he will - because of the substantial utility discount in the next period - again want to stop smoking "tomorrow," since the lost utility will then have a less dramatic effect on the function. Thus, with a clear conscience and the firm resolve that this will be his last, he smokes his cigarette. The next day, however, the preferences will have shifted so far that the future utility of stopping will not be sufficient to support the decision any more.

The quasi-hyperbolic model can also be easily calibrated. The general discount factor δ can be determined just like in the discounting model, with the difference that only alternatives that do not have outcomes in $t = 0$ are compared. If δ is known, the preference for immediate consumption β can then be identified by comparing alternatives for which $t = 0$ is also relevant. As usual, consistency checks should be performed to avoid mistakes.

We have seen that none of the axiom systems just introduced rely on axioms that are indisputably rational, and that therefore none can be regarded as generally superior to the other. You have to decide in each individual case whether you want to use one of the special models or the general additive model. To find the appropriate model, two basic procedures are possible.

First, clarify the implications of the chosen axioms by using examples (like the summer vs. winter vacation example above) to determine whether you think that the chosen axiom might provide a sound basis for rational decisions in the given situation. If you accept stationarity as a condition for rational behavior, you have identified the discounting model as the appropriate one. Similarly, if you accept impatience and invariance under stretching, you will choose the Harvey model. However, if you find all three axioms acceptable, you have gotten caught in the depths of axiomatic decision theory – the axioms are incompatible; in other words, your opinion is logically inconsistent. Taking these axioms into consideration of course means a willingness to deal with their logical premises. In practice, not every decision maker will be prepared to do this.

At this point, we would therefore like to illustrate again the notions of stretching and shifting a sequence of outcomes by providing an example. As a member of the city council, you are involved in a program to create new kindergarten places. Your committee's objective is to maximize the annual number of new kindergarten places.

Program a provides for the building of new kindergartens within the next five years, to be financed directly by the city. The annual budget allows for the creation of 650 new places per year from period one onwards; no places will be created in the construction period zero. Program b uses the same budget over the next five years as municipal aids for private enterprises that will create daycare facilities for their staff. This program will create 200 places in the current year, and, with increasing awareness and acceptance in the private sector, afterwards 200 places more each year than in the previous year. Programs a and b can be represented as the following sequences of outcomes:

$$a = (0, 650, 650, 650, 650)$$
$$b = (200, 400, 600, 800, 1.000)$$

After its commitment to the necessary funds, the city falls into dire financial straits and the money cannot be allocated as initially approved.

A shifting of the sequences of outcomes results if the city cancels all funding for, say, the next two years, and the project will be started with a delay of two years. The sequences of outcome $\alpha_2(a)$ and $\alpha_2(b)$ then read:

$$\alpha_2(a) = (0, 0, 0, 650, 650, 650, 650)$$
$$\alpha_2(b) = (0, 0, 200, 400, 600, 800, 1.000)$$

If after each year the project is interrupted for one year, the result is a stretching by a factor of 2. The reason for the interruption might be the city's need to evaluate for instance the experiences of the participating families and enterprises. Under this stretching the sequences would look like this:

$$\beta_2(a) = (0, 0, 0, 650, 0, 650, 0, 650, 0, 650)$$
$$\beta_2(b) = (0, 200, 0, 400, 0, 600, 0, 800, 0, 1.000)$$

Now try to decide which of the two original sequences of outcomes a and b you, as a council member, would prefer. Does your preference change after shifting or stretching? Maybe program b will at first seem more beneficial because if you manage to create 1,000 kindergarten places annually in five years, your election prospects will improve. However, after shifting - or even stretching - these profitable periods would be postponed so far into the future that a program creating 650 places fairly quickly would seem more favorable. Stretching or shifting tips your preference in favor of a: in this case, neither the discounting model nor the Harvey model will be able to display your preference. You will have to fall back on the general additive model, and probably also - on the basis of the above considerations - depart from the assumption that your period value function is identical for each point in time.

A second, more viable possibility for the choice of your model is to evaluate the models according to the results they produce. Such a procedure was chosen for instance in Chapter 9: here we rejected the expected value criterion as a basis for rational decisions under risk because the consequence from this criterion would be an infinite value for the St. Petersburg game. The discounting, the Harvey, and the quasi-hyperbolic discounting models only differ in the weights they assign to individual periods. To better understand how the discounting formula works, we need to compare the period weights in these models. To this purpose, one could either deduce general statements or simply consider an example like the following.

In this example, we have chosen a discount factor of $q = 1.05$ for the discounting model. For both the Harvey and the quasi-hyperbolic models, we have chosen parameters that assign the same weight to period 15 for all three models, with $\beta = 0.6$. This gives us $r = 0.264$ and $\delta = 0.9853$. Figure 11-3 shows the period weights of the models for the first thirty periods.

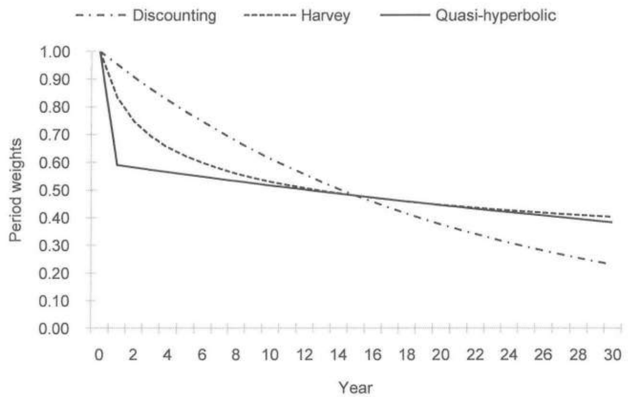


Figure 11-3: Period weights in the discounting, the Harvey and the quasi-hyperbolic discounting model

The chart clearly illustrates the differences between the models. The discounting model assigns significantly lower period weights to periods further in the future than the other two. Even when using different parameters, this effect still remains. In the quasi-hyperbolic discounting model, the preference for immediate consumption also stands out distinctly, expressed by the choice of parameter β.

The strong discounting of future consequences in the discounting model does not necessarily speak against the model. In many decisions, it is not necessary to consider a great number of periods, the strong discounting would not even come into effect in this case. Even for decisions reaching far into the future, e.g. regarding your health in old age or environmental decisions, it is quite possible that you appreciate the significance of the remote future according to the weights of the discounting model. On the other hand, you might find the comparatively weak weighting of the remote future not rationally acceptable while still wishing to hold on to the discounting model. In this case, you will have to reconsider your steps for determining the appropriate discounting factor, perhaps revise them and specify a smaller discounting factor. If, however, the reassessment of your discounting factor does not bring about a modification of the factor, you cannot regard the discounting model as a sound basis for a rational decision. You should then consider other models more closely or use the general model for your decision.

11.4 Evaluation of payment sequences

So far, we have derived different functions for the individual intertemporal value function of a decision maker. If the consequences are defined as payments, the capital market may transform these payments from one period to the next. Some

examples for transformation options the market offers are borrowing, putting money into your savings account or a fixed-term deposit. We will assume that for each period, there are transformation options that will allow us to move the payments relevant to the decision problem from one period to the next. Additionally, we will assume that for lending or borrowing from period $\tau - 1$ to τ, the relevant interest rate is i_τ. Based on these assumptions, any sequence of payments can unambiguously be transformed into an equivalent and simple payment sequence with payment occurring only in period zero. This variable in period zero is called capital value and can be computed in the following manner:

$$v(x) = \sum_{t=0}^{T} \frac{1}{\prod_{\tau=1}^{t}(1+i_\tau)} x_t. \qquad (11.10)$$

In this formula, we have assumed a specific period interest rate for i_τ for transactions between periods $\tau - 1$ and τ. If the interest rate for all period transitions is equal to i, Formula (11.10) simply becomes

$$v(x) = \sum_{t=0}^{T} \frac{1}{(1+i)^t} x_t. \qquad (11.11)$$

For the derivation of the capital value, no axiomatic considerations are necessary: the mere fact that options for transformations exist warrants the existence and the form of the capital value (for the derivation of the capital value using the Fisher separation, see Franke and Hax 2009). We can, however, interpret the capital value as an additive intertemporal value function and thereby determine which axioms a market with transformation options implicitly fulfills. The capital value implies linear period value functions, i.e. $v_t(x_t) = x_t$. You obtain the weight for period t by multiplying the first t interest rate factors $1 + i_\tau$ (Formula (11.10)). If all $i_\tau = i$, you will obtain period weights that correspond to the discounting model with discount factor $q = 1 + i$ (Formula (11.11)).

Let us use an example to demonstrate the connection between capital value and intertemporal value function. For simplicity's sake, let us suppose $T = 2$. The interest rate factors are $i_1 = 6\,\%$ and $i_2 = 7\,\%$. Table 11-3 shows two projects between which a decision maker has to choose.

Table 11-3: Two payment sequences a, b and their capital values

	$t=0$	$t=1$	$t=2$	$v(.)$
a	–€150	€87	€93	€14.08
b	–€140	€69	€102	€15.03

According to the interest rates given, and hence the implicit value function, alternative b is preferable. To illustrate this, let us consider the transformation of a into the sequence of payments with payments only in $t=0$. Here, you can use the payments from $t=1$ and $t=2$ to repay your loans; you just borrow enough money in $t=0$ to be able to use the €87 (€93) from $t=1$ ($t=2$) to cover repayment and returns. Considering the given interest rates, you can thus take out a loan of €82.08 (plus €82.00) in $t=0$, i.e., a sum total of €164.08. After subtracting the €150 payable in $t=0$, you get a remainder of €14.08. Applying the same method to b you get a remainder of €15.03, i.e., the capital value resulting from the transformation of the payment sequences is bigger for b than for a.

The capital value rests on the assumptions made concerning the options for the transformation of payments offered by the capital market. Identical debit and credit interest rates for the payments relevant to the decision problem may at first seem like unrealistic premises. For big corporations with access to the capital market, debit and credit interest rates however differ so slightly that the difference can be neglected. But even for small and medium-sized businesses, or private persons, these premises can frequently be accepted. If a decision maker is in debt, his debt will increase with negative payments and decrease with positive payments. In both cases, the interest rate (in this case, the debit interest rate) is relevant for the transformation of payments. Similarly, for decision makers with capital, usually only the credit interest rate is relevant.

As a rule, the problem of time preference for payments vanishes if values can be transferred from one period to the next. Generally speaking, transferability defines an external exchange rate between periods. This external exchange rate allows the decision maker to evaluate alternatives independently of his (internal) time preferences by choosing the alternative (payment sequence) with the maximum capital value (capital value dominance). The transferability (i.e. the existence of an external exchange rate) is not limited to payments. Futures markets, for instance, offer the possibility of transferring other goods while at the same time determining external exchange rates. Furthermore, transferability does not necessarily require the existence of markets. At this point, we will however not go into any further detail. The derivation of exact conditions for the replacement of internal time preferences by externally given exchange rates would indeed go beyond the scope of this textbook.

Questions and exercises

11.1

You have decided to use the discounting model in a decision situation in which you have to consider 25 periods. From an indifference statement for the first two periods you have derived the discounting factor to be 1.08. When finding out that by doing so, the weight for period 25 is less than 15% of the weight for period 0, you consider to either change the discounting factor or the model. You think the weight for period 25 should be two thirds of the weight for period 0.

(a) Which discounting factor would you have to choose?

(b) Which parameter would you have to choose for the Harvey model?
(c) How would you have to proceed if as a rational decision-maker you wanted to define the weight for period 25 as two thirds for period 0, and want to stick to the indifference statement for periods 1 and 2?

11.2

Think of three decision situations, e.g. from your private life, in which you have to consider time aspects.

(a) Do you think that the axiom of preference independence is fulfilled between the periods?
(b) Which of the three discussed models of this chapter—the additive model, the discounting model or the Harvey model—seems to be most appropriate for each situation?

11.3

In the discounting model, you can also determine the discounting factor q by knowing the half-value period h.

(a) How do you determine the discounting factor q from a given half-value period h?
(b) What are the advantages and disadvantages of defining the discounting factor from the half-value period instead of using indifference statements regarding periods 0 and 1?

11.4

You decided to use the Harvey model, but it is hard for you to define the half-value period. You can only state that h is between 60 and 75. You have to choose between two alternatives a and b. In nearly all periods, a has the same consequences as b. Only periods 0 and 10 have different consequences. From the following table, you can see the evaluated outcomes of a and b:

Period	a	b
0	1	0.8
10	0.5	0.8

Can you make a rational decision between a and b although you cannot determine the exact half-value period? If so, which? If not, can you at least make some statement?

11.5

Consider the sequence of (evaluated) outcomes $a=(1, 1, 0, 0, 0, ...)$. Which of the sequences b, c, d, e given below are constructed by shifting and/or stretching the sequence?

$b=(2,2,0,0,0,...)$, $c=(0,1,1,0,0,...)$, $d=(1,0,1,0,0,...)$ $e=(0,1,0,1,0,...)$

11.6

(a) John uses the Harvey model. He is indifferent between the (evaluated) sequences $A = (0, 2, 0, 0, \ldots)$ and $B = (0, 0, 3, 0, \ldots)$. Determine John's period weights $w_0, \ldots, w_3$.

(b) Name another sequence C of evaluated outcomes that makes John indifferent between A, B and C.

(c) Michael uses the exponential discounting model. Is it possible that Michael is also indifferent between the three sequences A, B and C (C is your answer in subtask b)? If yes, provide appropriate period weights for the exponential discounting model. If no, give informal arguments.

(d) Could your answer in subtask c) change if you provided a different sequence C in subtask b)?

11.7

(a) Is it possible to model time-consistent preferences in a quasi-hyperbolic discounting setting? If yes, how? If no, why not?

(b) How would the stationarity axiom have to be modified to yield the model of quasi-hyperbolic discounting?

11.8

An exponential discounter ($q = 1.02$) and a quasi-hyperbolic discounter ($\beta = 0.4$, $\delta = 0.98$) are facing a decision regarding their retirement savings: in period $t = 0$ (now) and $t = 1$ (next year), they can consume their income to generate a utility of 10 each period or they can start a payment plan for their retirement savings account. The resulting utility if they decide to go with the payment plan is equal to 5 per period. When they retire ($t = 40$), they have a utility of 25 if they start saving immediately ($t = 0$; Option 1) or 10 if they start saving next period ($t = 1$; Option 2). Not saving at all (Option 3) will result in a negative utility of -40 when they retire. Periods $t = 2$ to $t = 39$ do not influence the decision and shall not be considered further. See the table below for a summary:

Period	$t=0$	$t=1$	[...]	$t=40$
Option 1	5	5	[...]	25
Option 2	10	5	[...]	10
Option 3	10	10	[...]	–40

(a) For both the exponential and quasi-hyperbolic discounter, calculate the expected utility for each option. Which option would they choose?

(b) If he had to decide in a similar manner each period, how much would the hyperbolic discounter spend on his retirement savings? Why?

11.9

An impatient investor ($q > 1$) without access to a capital market (i.e. without options for temporal transformation) has €100 available to invest. He has the choice

between two alternatives with the following payoff profiles (the payoff in period $t = 3$ includes all liquidation proceeds of the investment):

Period	$t=0$	$t=1$	$t=2$	$t=3$
Investment 1	–€100	€80	€150	€300
Investment 2	–€100	€80	€140	€315

(a) Based on this information, can you already draw a conclusion which alternative the investor will prefer?

(b) Assuming that the payoff of investment 2 in period $t = 3$ equals €310 instead of €315. Can you draw a conclusion now? What if the time preference of the investor was given by $q \geq 1$?

(c) What information concerning the time preferences of the investor would you need if the payoff of investment 2 in period $t = 2$ was €150?

Chapter 12: Group decisions

12.0 Summary

1. In this chapter, we discuss group decisions based on objectives and expectations jointly set by the group members.
2. In decision making processes within groups, issues frequently appear which affect the rationality of the process.
3. Jointly structuring a decision problem creates a higher level of procedural rationality than voting on given alternatives after an unstructured discussion. This tends to make a problem more transparent and curb the influence of "illegitimate" interests.
4. If the group is not able to agree on specific values (like probabilities or the weighting of objectives), one can either aggregate the individual values mathematically or make the group agree on a certain interval, accepting to continue working with incomplete information.
5. An efficient possibility to aggregate probability estimations made by experts is given by so-called prediction markets. On these markets, experts receive monetary incentives to trade according to their probability predictions.
6. After the components of the decision problem have been determined, a solution can be found entirely analogously to the case of an individual decision.
7. The alternative to joint structuring is that each person states their own preferences concerning the alternatives. These individual preferences are then transformed into a group preference, either by aggregating the individual utility functions or by voting on the given alternatives.

12.1 Benefits and problems of group decisions

Very few important decisions are made by individuals on their own. The complexity of problems usually requires the collaboration of experts from different work fields such as marketing specialists, engineers and financial managers. Besides, in our cultural environment there is a common belief that the people affected by decisions should also participate in reaching them. People who participate in making decisions also tend to be more involved in the successful implementation of the solution. Many public and private corporations assign decision making to committees such as the board of directors, a faculty or a cabinet, either by law or by statute.

In this chapter we examine how the principles of prescriptive decision theory can be used in the field of group decisions.

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_12, © Springer-Verlag Berlin Heidelberg 2010

12.1.1 Adverse group effects

The potential advantages of group decisions over individual decisions are evident: more people can generate more ideas, they bring in more knowledge about facts and relationships and incorrect assessments of individuals may be compensated for. By exchanging and reviewing arguments, all participants get the chance to revise their judgments.

On the other hand, extensive socio-psychological research as well as daily experience show that in a group environment, certain mechanisms may interfere with - or even completely obstruct - the advantages obtained from the group decision making process. The extent to which these negative effects occur depends on the composition of the group on the one hand and the type of decision task on the other hand. Regarding the group composition, the following criteria are particularly important:

- The level of group cohesion;
- Differences in power and status among group members; and
- Conflicts of interest among the members.

Cohesion describes the solidarity of the group which is affected by the attractiveness of the group for the members. Although (or because) solidarity promotes striving for a consensual solution, an excess of cohesion is also one of the reasons for a phenomenon which Janis (1972) described as *Groupthink*. This is characterized by a strong striving for consensus, premature agreement without checking a sufficient amount of alternatives, a feeling of superior competence of the group and resistance to all counter-arguments. A variety of disasters which with hindsight seem absurd might be traced back to the Groupthink phenomenon, such as the Vietnam War, the Watergate scandal, John F. Kennedy's 1962 invasion of Cuba, the explosion of the space shuttle *Challenger* (Kleindorfer et al. 1993) and the collapse of the British bank Northern Rock. Regarding the last case, Kamau and Harorimana (2008) argue that the bank's management lacked objectivity, thought of its own business model as infallible, was prone to risk shifting, and remained isolated from contradictory influences.

Distinctive *differences in power and status* may give objectively unjustified preference to a solution favored by a dominant individual. Such a person will often claim more speaking time than others. Members of a lower status will usually avoid defending opposing positions, and may eventually - maybe even unconsciously - adapt their original opinion to that of a higher-ranked group member.

If *personal interests* of group members are affected by the upcoming decision, the assessment of the alternatives is almost unavoidably influenced by those interests. The tendency of individuals to share information which could be detrimental to their own interests naturally is weak. On the contrary, contributions to the discussion that exaggerate the advantages of the preferred solution and specifically disparage those of the competitive suggestions are more likely. The advantage that less informed group members profit from expert knowledge of others is then reversed: experts may take advantage of being leaders in knowledge.

Moreover, group discussions may *put excessive strain on the cognitive capacity of the group's member'*. Concentrating on the course of the debate may leave little room for personal reflection. Particularly under time pressure, creativity cannot be sufficiently developed. The more often an argument is repeated, the greater its influence. During the course of the discussion, arguments and suggestions are frequently forgotten, and those that are contributed late in the discussion often prove to be decisive. Cognitive restrictions are also a reason for a lack of alternatives and aims proposed for discussion; the discussion will often prematurely focus on only one or two alternatives.

12.1.2 Potential remedies

The insight into the defects of group decision making has led to a variety of concepts on structured group decisions. In Chapter 4, we already pointed out *Brainstorming* and the *Nominal Group Technique* which reduce creative blocks and the influence of status differences.

A number of recommendations against the *Groupthink* phenomenon have been made by Janis and others:

- It has to be ensured that all objectives are explicitly mentioned;
- Members should be encouraged to express doubts about assessments;
- High status individuals should not present their preferences at the beginning;
- The group should be split up to produce alternative solutions;
- For every suggestion, an *advocatus diaboli* (devil's advocate) should represent an opposite opinion;
- Before implementing a decision, each group member should voice their doubts about the chosen alternative.

Technical support can be provided by a computerized *Group Decision Support System* (GDSS). Here, every participant uses a computer with internet access which communicates with a server. Anonymity reduces the influence of status differences and the dangers of conformity. Ideas can be brought forward at any time without having to wait for speaking permission. The danger of forgetting is removed; everybody has access to a summary of the state of the discussion at any time and can print out a set of minutes. A decisive advantage is the fact that simultaneous presence of all participants at the same place at the same time is no longer required.

Furthermore, GDSS can support the decision making process with suitable software. First, the system can be connected to databases and give everyone access to the same information. Second, a model bank can provide programs developed for the support of individual decisions which can be used by the group as well, e.g. programs for the joint creation of flow charts or decision trees. Third, software can aggregate individual opinions into a group opinion, recognize dominated alternatives and, if necessary, suggest a solution for the decision problem. Geibel (1993), and Davey and Olson (1998) provide a summary of different specifications of GDSS.

Chen et al. (2007) have developed a web-based GDSS and have tested it in various empirical studies. They find that users of their software achieve a better team performance than users of the control group without GDSS support. Modern office software often offers basic GDSS functions, take for instance the SharePoint Workspace integrated in Microsoft Office 2010.

12.2 Joint structuring of the decision problem

Beyond the recommendations that prescriptive decision theory offers to individuals to form and process their aims and expectations, to what extent can the theory also help achieve a higher level of rationality of *group* decisions? The decomposition of the decision problem into several modules – aims, alternatives, expectations, and effect models – should provide a basis for structuring the group process. There is no separation of this kind in the unstructured group discussions which all of us know. Suggestions, assertions, assessments (which implicate aims) are thrown into the debate in random order. It is hardly possible to guarantee completeness, non-overlapping, and fundamentality of the criteria in such a discussion; the search for alternatives often remains inadequate and focuses on a few solutions close to the status quo. Expectations about uncertain developments are not quantified but expressed in extremely vague colloquial language. On this basis, every individual forms an opinion about the best possible suggestion, and in the end there is a decision by vote.

In contrast to this, the decompositional approach offers the opportunity of higher procedural rationality. Each module is discussed individually by the whole group; the creativity and the knowledge of all members can thus be used for *every* single component. There is no need for the individuals to solve the decision problem on their own, possibly with insufficient know-how and insufficient structuring. Only *one* person needs to have the knowledge about choice-theoretic methods in order to serve as a neutral moderator of the group discussions.

Just like an individual, a group generally cannot model the modules in a linear run either but will occasionally have to revise already processed modules. In group decisions, you have to ensure that all members are always working on the same module at the same time. This requires setting up an agenda. To preserve the structure of the process, it is the moderator's task to maintain a certain discussion discipline.

The explicit modeling of all components guarantees the greatest possible transparency and traceability of the decision basis. Conflicts can be easily identified because there is no argument about complex alternatives but about clearly defined facts – such as the probability of an event or the weighting assigned to an objective.

The transparency of the decision basis facilitates identifying reasons for success or failure *ex post*. Failures can, for example, be explained by the fact that objectives important today might not have been taken into consideration at the time of the decision or that events which occurred might have been considered very improbable at the time. Those insights are instructive for future decisions and can help to avoid unjustified blame for failures.

Furthermore, the possibility of pursuing "illegitimate" objectives is constrained. We define illegitimate objectives as objectives which contradict the mandate given to the group. For example, if a human resources committee has the task of choosing the most suitable applicant, it would not be compatible with their mandate if a commission member championed one applicant only because he or she hopes for personal benefits from that applicant. Arguments for or against an alternative can often be assessed only with difficulty because they concern merely one or a few aspects; this creates the opportunity to argue for a point that is driven by purely egotistical motives. An egotistically motivated opinion can obviously also be introduced into the discussion about value functions or about probability distributions; however, the likelihood of discovering information distorted by individual interest is much greater there. Moreover, extreme assessments not shared by the other members can be adapted to the group average in a process of aggregation.

The conditions under which a group accepts the described procedure, i.e. the joint modular structuring of the problem and the aggregation of individual judgment within these modules largely depend on both the composition of the group and the type of the decision problem. Following March and Simon (1993, p. 149 ff.) we distinguish between four decision situations:

1. *Problem solving*. The group shares common objectives and looks for the best solution. The task is to collect relevant information and to generate new alternatives.
2. *Persuasion*. Apart from shared there are also differing objectives. The differing objectives, however, tend to be of an instrumental nature while the corresponding fundamental objectives are still shared by everyone. These differences can therefore be overcome by reasoning.
3. *Bargaining*. Conflicting objectives between participants are based on diverging interests and cannot be removed by persuasion. The group will not look for the best but - at most - for a "fair" solution. Threats, wrong assertions and "horse trading" are parts of the rules of the game within the group.
4. *Politics*. There are open conflicts of objectives similar to those in situation 3, however the negotiation situation is not limited to a certain field and the parties are looking for external allies.

Obviously, only the first two situations (which March and Simon identify as "analytical" processes) are suitable for the decision-theoretic procedure. The two other situations are subjects of bargaining theory and game theory. Bargaining theory is not based on the assumption of a group with common aims but of negotiating parties that want to maximize their own benefits (on bargaining theory, see Roth 1979 and Raiffa et al. 2003). A better solution than the status quo can be achieved only if the opponents agree on an alternative. The theory examines which alternatives are rationally to be taken into consideration for an agreement. Nash (1950) derived a clear negotiation solution from some axioms. The solution is the alternative

which maximizes the mathematical product of the benefit augmentations (compared to the status quo) of all negotiating parties.

In game theory (von Neumann and Morgenstern 1947), decisions are examined in which multiple decision makers with conflicting interests are in opposition to each other. Here, however, each decision maker decides autonomously; the participants do not act as a group nor is the result negotiated. The arms race of the super powers after the Second World War is a typical example of game theoretical situations. We refer the interested reader to Rasmussen (2006), Berninghaus et al. (2006), and Holler and Illing (2008).

Even with a basic agreement about the aims in place, it is possible that a common structuring does not take place or is not carried through. The "group" may simply be too big (for example a parliament), may have no time for analytical processes or may not know or trust decision-theoretic methods.

The antipole of a joint structured analysis is the procedure that all group members – perhaps after a group discussion – form their own preferences regarding the alternatives. The group decision is then obtained by transforming the individual preferences into a group preference. This is usually done by voting on the alternatives; however, aggregating individual value or utility functions to a group function is equally possible (Keeney and Raiffa 1976, Dyer and Sarin 1979, French 1988). The group value function or group utility function would then be used for the assessment of the alternatives.

In this chapter, we assume that the group is ready to develop all modules of the decision model within the group process. We concentrate on the generation of a common system of objectives, the joint assessment of certain alternatives and the generation of joint probability distributions.

12.3 Generation of a common system of objectives

A common system of objectives is created simply by uniting all individual objectives. We assume that the group has certain fundamental objectives in common. Initial conflicts are to be examined with regard to whether or not they can be resolved by recourse to these fundamental objectives. For illustration, assume a married couple plans a vacation together; the husband wants to stay as close to the beach as possible, while the wife would prefer to stay as far away as possible from the coast. Are these really the fundamental objectives, or are other, more fundamental objectives behind these? Maybe the partners discover that their real, shared objectives are quietness and quick access to the water.

Another problem appears when a member does not share a certain objective or even disapproves of it. For example, a human resources committee might define the religious orientation of applicants as a criterion, which some members could disapprove of. Of course, the opponents could be advised that they could assign a zero weight to this objective. However, this would not prevent the controversial objective from gaining a positive weight when the group later aggregates the weights of the objectives. If the group fails to reach an agreement on the system of objectives by reasoning, the members should agree to evaluate the alternatives before the final decision both with and without the controversial objective in order to identify whether this objective affects the choice.

A conflict about the weighting of objectives does not exclude a common system of objectives. For instance, as opposed to an employee representative, a shareholder's representative on the supervisory board may give more weight to profit interests than to employee interests. If both want to maximize profit and preserve as many jobs as possible, they can easily agree on a common system of objectives; the conflict will however manifest when the weighting factors are to be determined.

When combining the individual aims, normally there will be much overlap and double-counting which have to be removed. Different wordings may have the same meaning; identical wordings may have different meanings.

The resulting hierarchy of objectives must fulfill the same requirements as that for individual decisions (Chapter 6). However, an unusual feature of the group decision is that, despite the agreement on the same objectives, there is no guarantee that the independence of preferences and utilities (desirable with regard to the decomposition of the assessment) applies to everybody in the same way. Commission member A could base the trade-off between exam grades and personal impression on professional experience: the more working years have passed since graduation, the less the exam grades count in proportion to personal impression. Member B however might see independence of preference between these attributes. If this discrepancy cannot be overcome by persuasion, there should be an attempt to redefine the attributes in order to achieve mutual independence of preference for all members.

In an extensive study, Keeney et al. (1984) gathered data about systems of objectives for the energy policy of the Federal Republic of Germany from several organizations and aggregated them to a common system of objectives. With regard to the procedure, we refer to this source.

12.4 Generation of group value functions

12.4.1 Aggregating individual single value functions

Aggregating individual evaluations inevitably leads to the problem of interpersonal utility comparison. There is no absolute utility scale by which individual utility could be measured and compared. Let's assume that two business people X and Y want to meet for a business lunch at a restaurant. X evaluates restaurant a with 0, restaurant b with 0.7, and restaurant c with 1. His partner Y however evaluates a with 1, b with 0.5 and c with 0 points. Adding up the values results in the choice of b. Compared to restaurant a, this equals a utility increase of 0.7 for X and a utility decrease of 0.5 for Y. Altogether, is that a gain in net utility? We do not know because we cannot find out whether X's utility increase of 0.7 is more important for X than Y's utility loss of 0.5 is for Y.

In principle, the problem of the interpersonal utility comparison is unsolvable; such utility comparisons however nevertheless implicitly influence almost all decisions. If we do not want to wearily reject the possibility of rational decisions in groups, committees or parliaments, we have to assume that we can add the utility of X to the utility of Y. Even with purely individual decisions we have a similar problem: we make a choice, the consequences of which we will have to live with for decades, but we do not know whether today's utility will be comparable with tomorrow's.

We will now assume that a joint system of fundamental objectives shared by all members has been worked out. An attribute is allocated to every objective at the lowest level of the hierarchy of objectives. The group members have also agreed on the relevant range of outcomes for each attribute. In the following, we will concentrate on decisions under certainty and on the additive multi-attributive valuation model. The group is confronted with the task of generating both a measurable common (single-) value function for every attribute as well as a weighting system.

In the beginning, every group member j has an incomplete or complete idea of his individual value function $v_{rj}(x_r)$ with regard to every attribute r. The (information regarding the) value function is complete if the member can assign an unambiguous evaluation $v_{rj}(a_r)$ to each alternative a regarding the attribute r; we then speak of unambiguous individual evaluations. The (information regarding the) value function is incomplete if the group member can only make a statement about the evaluation interval, i.e. indicate a lower limit $v_{rj}^U(a_r)$ and an upper limit $v_{rj}^O(a_r)$ for the evaluation. In this case, we speak of individual evaluation intervals. Since we do not make any assumptions about the size of the interval, this initial situation is always given.

Unambiguous individual evaluations (complete individual information)

First, we want to consider the case where all group members give an unambiguous evaluation of alternative a with regard to attribute r. Obviously, if all unambiguous evaluations are identical, they can be adopted directly as the group evaluation; otherwise, there is a dissent which might not be easily resolved.

Solutions

In terms of a rational decision making process, there should at first be an attempt to resolve the dissent by reasoning. In particular, individuals with extremely divergent evaluations will be expected to provide convincing arguments. Baucells and Sarin (2003) show that, under certain circumstances, it may be sufficient in bilateral discussions to reach a compromise for single parameters from which the group preference can then be derived. If this attempt fails, there are two possibilities: mathematical aggregation or retention of an incomplete evaluation. A mathematical aggregation will usually be an average accumulation. This could be improved by weighting the individual evaluations in order to take into account the members' differences in competence. Let g_{rj} be the weight which member j is assigned regarding the attribute r, then the evaluation result is

$$v_r(a_r) = \sum_j g_{rj} v_{rj}(a_r). \tag{12.1}$$

The literature also suggests some (in part significantly more complicated) aggregation mechanisms to minimize the discrepancies between individual evaluations and the group consensus (González-Pachón and Romero 2006).

The other possibility is to leave the evaluation open until a later dominance analysis (see Section 12.5) has established whether the discrepancies in evaluation have any effect on the decision at all. Mathematical aggregation may have the advantage of finding a clear evaluation; its disadvantage however is that it can lead to a decision which is not compatible with the evaluations of all members. In terms of the approach presented here, this is not desirable. We would therefore recommend an immediate average accumulation (without prior dominance considerations) only if the differences between the individual evaluations are insignificant and if all participants approve.

Individual evaluation intervals (incomplete individual information)

If some group members can give their evaluation only within certain intervals, the situation becomes even more complex. Suppose the evaluation of alternative a regarding the attribute r made by person j lies within an interval between $v_{rj}{}^U(a_r)$ and $v_{rj}{}^O(a_r)$. First, it is necessary to clarify how such a statement is to be interpreted. In the discussions above (objective weights, probabilities, and value and utility functions) we had always assumed that the decision maker has difficulty giving a more precise assessment. On the basis of dominance considerations we therefore checked whether it is necessary at all to ask for more accurate evaluations; this approach can also be transferred to situations of group decisions. Let us therefore consider a comprehensive evaluation interval

$$\left[\min_j v_{rj}^U(a_r); \max_j v_{rj}^O(a_r)\right] \tag{12.2}$$

and examine via dominance tests if taking all possible evaluations of all group members into account would lead to the same decision. Obviously, in the case of strongly diverging evaluations among a great number of group members, this approach would result in very broad intervals making it highly unlikely to detect dominating alternatives. Should, however, only a few group members be responsible for the overall width of the interval because they have given intervals widely differing from the rest of the group, the reasons for these divergences can be systematically analyzed. Exchanging arguments and making slight changes to the system of objectives might already lead to a much narrower group evaluation interval and increase the chances of detecting dominant alternatives. In such a scenario of incomplete individual information, a formal aggregation mechanism leading to an exact group evaluation should be applied as carefully as in the case of complete individual information. Such an aggregation mechanism could, for instance, condense the individual intervals to their corresponding mean, which then again could be averaged over all group members. Within this procedure, a weighting could be given according to the group members' competence, just as in the case of unambiguous individual evaluations. However, this procedure raises new issues since the group now has to decide on a common weighting scheme. Unfortunately, no satisfying suggestions have been made to solve this problem - for this, see also the discussion in Han and Ahn (2005), who suggest - if applicable - to only determine intervals and then check with dominance tests whether the exact choice of weights plays a role at all. Alternatively, the average individual evaluations could be weighted by the inverse width of the individual intervals to incorporate the uncertainty regarding the individual evaluation into the group evaluation.

Since until now we have always assumed that the group members are willing to jointly find an optimal evaluation for the whole group, there is a further possibility to interpret the statement of evaluation intervals. The group member could wish to express that it approves of every group evaluation that falls within this interval. This perspective would require a very different conceptual treatment - broad intervals would be no longer a problem but would instead make it easier to determine the optimal alternative for the group. In this case, the individual assessments are called *compatible* if the minimum of the upper individual value limits is greater than or equal to the maximum of the lower individual value limits, so that

$$\min_j v_{rj}^O(a_r) \geq \max_j v_{rj}^U(a_r) \tag{12.3}$$

applies. For people able to give unambiguous evaluations, lower and upper limits coincide. Figure 12-1 shows compatible and incompatible evaluations.

If all members make incomplete evaluations and if these are compatible with each other according to Formula (12.3) and Figure 12-1(a), an evaluation margin remains between the minimal upper limit and the maximal lower limit, provided they do not coincide. The group then faces the "luxury problem" of picking one group evaluation from the set of evaluations which are acceptable for all group members. Under certain circumstances, it might be reasonable to retain the interval of acceptable evaluations to be able to later check with a dominance (or sensitivity) analysis whether the specific choice of evaluation actually plays a role in the decision. If this is not the case, the specific evaluation needs not be considered further. In addition, when applying this interpretation of evaluation intervals the method of simply condensing the interval to a point (e.g. the average) is less of a problem. Since the thusly generated group evaluation has the guaranteed property that it lies within the interval of acceptable evaluations for every group member, the procedure will hardly cause acceptance problems.

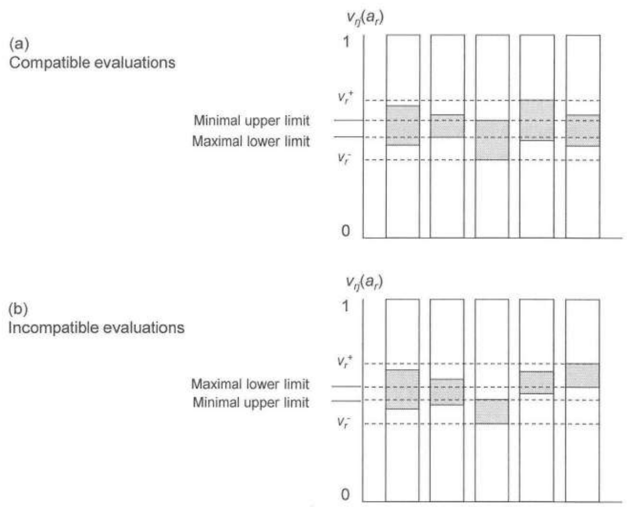


Figure 12-1: Compatible and incompatible individual evaluations

Should there be no overlap area indicating the incompatibility of the individual evaluations as in Figure 12-I(b), there is still the possibility of proceeding - in the same manner as in the case of unambiguous evaluations - with an interval of group evaluations hoping to identify the optimal alternative by dominance considerations. The group's evaluation interval would have to share a non-empty intersection with each individual evaluation interval and would at the same time have to be narrow. The latter makes it easier to identify dominance relationships. It is easy to see (for instance by looking at Figure 12-I(b)) that an interval must have the following form in order to meet these criteria:

$$\left[\min_j v_{rj}^O(a_r); \max_j v_{rj}^U(a_r)\right] \quad (12.4)$$

12.4.2 Generating common attribute weights

Suppose that a group wants to determine the attribute weights w_r according to one of the methods described in Chapter 6. According to the swing method, for example, m hypothetical alternatives are formed showing the best outcomes in exactly one attribute and the worst outcomes in all the others. The group then agrees on the preferences regarding these hypothetical alternatives and in the next step agrees on their evaluation on a cardinal scale. If this is not even successful for intervals, each group member should evaluate their own attribute weights using one of the aforementioned methods. Here again complete or incomplete information about the individual weights might result. At this point, we would like to remind you of the considerations in Section 6.5.3 in particular. Attribute weights reflect exchange relations between attribute outcomes. Therefore, incomplete information about attribute weights can only be the result of somebody not being able to give an exact weights ratio w_i / w_j but only an interval. As we have seen in Section 6.5.3, it is helpful to replace the common standardization technique (sum of weights must be equal to 1) by an alternative form ($w_l = 1$) in such a situation. The other weight intervals can then be determined without any further restrictions so that the exchange relations with attribute 1 are correctly reflected. At this point, we do not wish to rehash all the interpretations and scenarios we discussed in Section 12.4.1 because we have discussed them already with respect to utility functions. But since we are introducing a new perspective here, we would just like to mention again the case in which a group member interprets incomplete information to mean that any attribute weight within his stated interval is acceptable. Hence, it would (analogously to the utility functions) hold that the individual attribute weights r ($r > 1$) are compatible with each other if

$$\min_j w_{rj}^O \geq \max_j w_{rj}^U \quad (12.5)$$

i.e. if the lowest upper limit for w_r mentioned by any member is at least as big as the highest lower limit stated.

As in the case of individual attribute-wise evaluations, you can either use mathematical aggregation to generate a clear weighting scheme or retain the incomplete information. If the group members have produced unambiguous individual weightings and if these differ only slightly from each other, it will be easy to agree on a mathematical average. However, if there are considerable differences among the members and/or if single members have expressed uncertainty about their evaluations by giving their weights considerably wide intervals, mathematical aggregation should not be first choice. Instead, dominance tests should then first show whether or not alternatives can be eliminated despite the incompleteness of the weights.

12.5 Dominance tests

We have presented dominance tests under incomplete information about weights in Chapter 6 and for probabilities in Chapter 10. We have pointed out their significance, i.e. the fact that decision makers and experts need not be asked to provide excessively detailed information – in which they have only little confidence – to be able to reduce the number of alternatives and even find the best alternative. Dominance tests become increasingly important for group decisions since both individual indeterminacy and interpersonally different evaluations can be taken into account here. Instead of a problematic mathematical aggregation, it can first be checked whether the group members' evaluations allow us to identify dominated alternatives – alternatives that are, in the eyes of *all* group members, inferior to another specific alternative.

As we have seen in Section 12.4, dominance tests can be conducted on the basis of different interpretations of incomplete information and with different aims. The most extreme case arises if the total spread of individual evaluations and attribute weights is taken into account. If in this case one alternative dominates another, this result rests on very solid foundations because every group member's complete evaluation interval would lead to the same decision as the group decision. However, this kind of test is not equivalent to a series of dominance tests on an individual basis. Even if the same alternative for all group members proves to be dominant with respect to the individual evaluations, a dominance test need not necessarily identify the dominance of this alternative for the total spread of the individual evaluations. A simple example will clarify this phenomenon:

Suppose two decision makers comparing alternatives $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ have identical individual evaluations leading to the following value differences: $v(a_1) - v(b_1) = -1$, $v(a_2) - v(b_2) = +1$ und $v(a_3) - v(b_3) = +1$. However, their attribute weights differ slightly. Decision maker 1 clearly defines weights $w_1 = 1$ und $w_2 = 0.6$, but expresses uncertainty for w_3 assigning the evaluation interval [0.45, 0.6]. In contrast, decision maker 2 is certain with respect to $w_1 = 1$ and $w_3 = 0.6$ but states an evaluation interval for w_2 of [0.45, 0.6]. It is easy to show that for both decision makers every possible weighting scheme leads to a higher valuation of alternative a, i.e. there is a dominance relation of a over b. But if you consider the total spread of weights given by $w_1 = 1$, $w_2 \in [0.45, 0.6]$, and $w_3 \in [0.45, 0.6]$ it is not possible to detect the dominance of a over b. Setting the weights to $w_1 = 1$, $w_2 = 0.45$, and $w_3 = 0.45$ makes alternative b preferable.

As the example shows, the consideration of the total spread and the requirement that the same alternative is always preferred are unnecessarily strict requirements. The dominance test would also consider combinations of evaluations and attribute weights, the variance of which could represent different decision makers and would therefore not reflect the decision uncertainty of an individual. This weakness in detecting individual dominances is the price which has to be paid for the simplicity of aggregation (unifying the intervals), interpretation (group interval), and dominance test (only one instead of many tests).

The other extreme is a dominance test in which only the intersection of the individual evaluation intervals is considered. This kind of test can become necessary if the group members approve of all evaluations that lie within their own stated intervals (please note that this is a completely different starting assumption). If this leads to various combinations of evaluations and attribute weights that the group could chose (as they are acceptable for all group members), it should be checked whether the specific choice affects the decision at all. The intervals are much narrower here than in the case discussed before so that the detection of dominant alternatives is much more likely. The approach is similar in application but conceptually entirely different if there is no overlap, i.e. if the individual evaluations are incompatible. Here again the maximal lower limit and the minimal upper limit of the individual evaluations are relevant for the group evaluation, but now the minimal upper limit lies below the maximal lower limit. If dominance can be shown for the group intervals thus specified, it follows that the group decision is compatible with at least one of the group members' possible evaluations. The extent to which this dominance test differs conceptually from the test discussed before becomes obvious by the fact that if dominance relations cannot be detected, it is necessary to make the individual evaluation intervals "more compatible". In the dominance test considered first, the intersection of individual evaluation intervals was too large to allow for a clear decision based on dominance relations. In that case the specification of "less compatible" intervals will reduce that problem.

12.5.1 Considering the complete range of evaluations

In the case of an additive multi-attributive evaluation under certainty, an alternative a dominates an alternative b if

$$\min[v(a)-v(b)] = \min\sum_r w_r[v_r(a)-v_r(b)] > 0\,. \tag{12.6}$$

Intervals are permitted for both the weights (excluding w_1) and for the evaluations of the alternatives regarding the individual attributes. The lowest value given by any group member to an alternative on the attribute r is denoted by $v_r^-(\cdot)$, the highest value by $v_r^+(\cdot)$. Furthermore, let w_r^- with $r > 1$ be the lowest weight of the attribute r mentioned by any group member, and w_r^+ be the highest one. The minimization task is then

$$\text{Minimize} \quad \sum_r w_r[v_r^-(a_r)-v_r^+(b_r)] \tag{12.7}$$

regarding the weights, subject to

$$w_r^- \le w_r \le w_r^+ \quad \text{for all}\ r>1$$

and $w_1 = 1$.

Dominance of a over b is given if the minimum is positive.

If there is no dominance of a over b, the dominance of b over a is tested for by considering the same constraints and solving the problem

$$\text{Maximize} \quad \sum_r w_r[v_r^+(a_r)-v_r^-(b_r)]\,. \tag{12.8}$$

If the maximum is less than zero, a is dominated by b.

As particularized in Section 6.5.3, the solution to these optimization problems can also easily found by hand. For instance, for the minimization problem in (12.6) one would need to set the weights w_r of all attributes r with a negative value difference $[v_r^-(a_r)-v_r^+(b_r)]$ to their upper limit. In case of a positive value difference, the minimal allowed weight will be assigned. In case the value difference is zero, the choice of weight is irrelevant for the dominance test. This procedure ensures that the expression in (12.6) is minimized while taking the relevant constraints into account.

If dominance can be identified using this procedure, it will apply to every single group member in particular, i.e. each participant would find the same dominance in his personal valuation model.

12.5.2 Reducing the range of evaluation intervals

To reduce the range of variance, we will only consider the interval between the highest lower value limit and the lowest upper value limit designated by any member. In the field of compatible evaluations, the maximal lower value limit forms the lower end of the interval and the minimal upper value limit forms the upper end. Accordingly, the reverse is the case for incompatible evaluations. The reasons for which a dominance analysis over such intervals may be necessary have already been discussed in depth.

12.5.3 Example

Two individuals are to evaluate two alternatives a and b with regard to the three attributes X_1, X_2 and X_3. Table 12-1 contains their incomplete value and weighting statements. To enhance the legibility of the table, we have already included the normalization requirement $w_1 = 1$ by choosing appropriate upper and lower limits.

Table 12-1: Incomplete information for evaluations and attribute weights

	X_1		X_2		X_3	
	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
Person 1						
a	0.20	0.35	0.55	0.65	0.75	0.85
b	0.15	0.25	0.70	0.75	0.40	0.50
Weight	1	1	1.20	1.50	0.60	1.20
Person 2						
A	0.15	0.30	0.60	0.70	0.80	0.90
B	0.20	0.25	0.55	0.70	0.45	0.55
Weight	1	1	0.90	1.20	0.70	1.00

We first carry out a dominance test over the whole range of evaluations. The resulting ranges of evaluations and attribute weights of the group are illustrated in Table 12-2. Note, for example, that for the evaluation of a on attribute 3, the whole interval between 0.75 and 0.9 must be considered as well as the interval between 0.9 and 1.5 for weight w_2.

Table 12-2: Incomplete information for the total range of evaluations and attribute weights

	X_1		X_2		X_3	
	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
Group						
A	0.15	0.35	0.55	0.70	0.75	0.90
B	0.15	0.25	0.55	0.75	0.40	0.55
Weight	1	1	0.90	1.50	0.60	1.20

In order to test for dominance of a over b, the following problem needs to be solved:

$$\text{Minimize } v(a)-v(b)$$
$$=w_1(0.15-0.25)+w_2(0.55-0.75)+w_3(0.75-0.55)$$
$$=-0.1w_1-0.2w_2+0.2w_3.$$

Setting w_2 to its upper limit and w_3 to its lower limit leads to the minimum. Applying these weights gives us

$$\min[v(a)-v(b)]=-0.1\cdot 1-0.2\cdot 1.5+0.2\cdot 0.6=-0.28.$$

Since the minimum is negative, there is no dominance of a over b. The next step is to test for dominance of b over a.

$$\text{Maximize } v(a)-v(b)$$
$$=w_1(0.35-0.15)+w_2(0.70-0.55)+w_3(0.90-0.40)$$
$$=0.2w_1+0.15w_2+0.5w_3.$$

Since all coefficients are positive, a positive maximum can obviously be achieved even without setting the weights to their upper limits. In order to further practice the procedure, we however do so anyway and obtain the maximal value difference for the following weight distribution: $w_1 = 1$, $w_2 = 1.5$, and $w_3 = 1.2$, which yields

$$\max[v(a)-v(b)]=0.2\cdot 1+0.15\cdot 1.5+0.5\cdot 1.2=1.025.$$

Since the maximum is positive, a is not dominated by b; we cannot eliminate either of the alternatives.

Therefore, in a second dominance analysis, we reduce the variance ranges of all evaluations to the intersecting intervals. In this example, all individual evaluations (of evaluations and attribute weights) are compatible so that all intersections are non-empty sets. The specific intervals of evaluations and attribute weights valid for the group are given in Table 12-3.

Table 12-3: Incomplete information for intersecting intervals

	X_1		X_2		X_3	
	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
Group						
A	0.20	0.30	0.60	0.65	0.80	0.85
B	0.20	0.25	0.7	0.7	0.45	0.50
Weight	1	1	0.90	1.20	0.60	1.00

With these substantially limited intervals, the chances of dominance have improved considerably. We minimize:

$$v(a)-v(b)$$
$$=w_1(0.20-0.25)+w_2(0.60-0.70)+w_3(0.80-0.50)$$
$$=-0.05w_1-0.10w_2+0.30w_3.$$

We obtain the minimum for the weighting $w_1 = 1$, $w_2 = 1.2$ and $w_3 = 0.6$. It is

$$\min[v(a)-v(b)]=-0.05\cdot 1-0.1\cdot 1.2+0.3\cdot 0.6=0.01.$$

This minimum is positive. Alternative a has a higher value even for its least favorable permitted choice of individual evaluations and attribute weights and therefore dominates b.

If no adequate number of dominances can be found by these means, the group members' information input should be improved to adjust the individual evaluation and weighting intervals in such a way as to produce narrower group intervals for the dominance test.

We emphasize once again that in principle we regard it as more worthwhile to find a solution by dominance tests which are based on jointly shared evaluations of all group members, rather than to force a solution by mathematical aggregation over possibly a wide disparity of evaluations.

12.6 Generating joint probability evaluations

Differing probabilities can - just like evaluations - be mathematically aggregated or serve as input for dominance tests. Simultaneous incompleteness of evaluation and probability information however is difficult to handle. We will here focus on the mathematical aggregation, nevertheless pointing out that we view this approach as very problematic if there are large discrepancies between the individual estimates based on fairly different effect models. If an engineer estimates the success probability of a new product at 0.9 for technical reasons and a marketing expert estimates it at 0.1 from the perspective of customer needs, it makes little sense to agree on 0.5. In this case, the arguments of both parties should be first combined in an effect model.

12.6.1 Simple aggregation of individual estimations

To determine probabilities of discrete events or conditions, von Winterfeldt and Edwards (1986, p. 136), suggest simple average accumulation as an efficient aggregation mechanism, as shown in Table 12-4 for three group members and four conditions.

Table 12-4: Aggregation of probabilities

	s_1	s_2	s_3	s_4	Sum
M_1	0.20	0.32	0.28	0.20	1
M_2	0.15	0.40	0.35	0.10	1
M_3	0.13	0.36	0.40	0.11	1
Group	0.16	0.36	0.34	0.14	1

For continuous event variables, the probability density function of the group can also be determined by a simple weighted addition of the individual probability density functions. For n group members, every individual probability density function is weighted by $1/n$ unless different weightings of the persons according to their differences in knowledge are agreed upon.

For example, let us assume three managers would have to estimate the costs of a building project. Manager A assumes a rectangle distribution in interval [480, 560] thousand €, manager B a rectangle distribution in interval [500, 600] thousand €, and manager C a triangular distribution in interval [520, 580] thousand € with the densest value at 540. The three probability density functions hence are

A:	$f_A(x)=0.0125$	for $480\le x\le 560$
B:	$f_B(x)=0.01$	for $500\le x\le 600$
C:	$f_C(x)=(x-520)/600$	for $520\le x\le 540$
	$f_C(x)=1/30-(x-540)/1{,}200.$	for $540\le x\le 580$

The outcome of this is the aggregated probability density function $f(x)$

$=0.0125/3$	for $480\le x\le 500$
$=0.0225/3$	for $500<x\le 520$
$=(0.0225+(x-520)/600)/3$	for $520<x\le 540$
$=(0.0225+1/30-(x-540)/1{,}200)/3$	for $540<x\le 560$
$=(0.01+1/30-(x-540)/1{,}200)/3$	for $560<x\le 580$
$=0.01/3.$	for $580<x\le 600$

The distributions are represented in Figure 12-2.

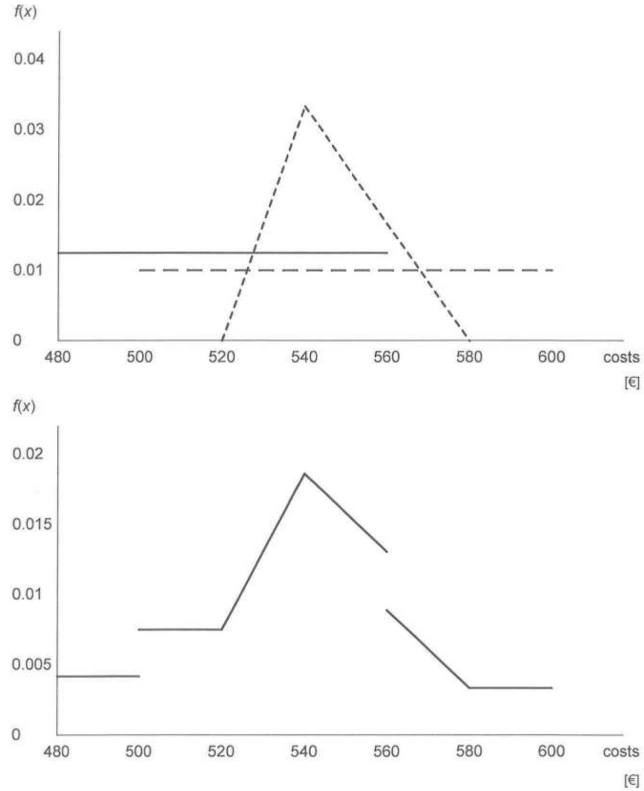


Figure 12-2: Aggregating the three probability density functions

Instead of probability density functions, distribution functions can also be used. Let us assume that two group members have given five supporting points of a distribution function for the future price of a commodity as shown in Table 12-5. For practical reasons, the same percentage points of the distributions should be used for the aggregation. An aggregated estimation results from simple averaging.

Table 12-5: Aggregating distribution functions

	Lower limit	25%	50%	75%	Upper limit
Person 1	13.00	16.50	18.00	19.00	22.00
Person 2	10.00	13.00	15.00	17.00	20.00
Average	11.50	14.75	16.50	18.00	21.00

It is of course possible to allocate different weights to the estimates of the members, provided the group agrees that this is how the individual's different levels of competence should be accounted for.

In addition to the simple average determination, von Winterfeldt and Edwards (1986, pp. 134-136) as well as Ferrell (1985) mention further methods for the generation of joint probability assessments which we do not want to discuss in detail. They all have their pros and cons which need to be balanced when choosing an aggregation method. The specific decision situation often has to be taken into account when making this choice. For example, when determining probabilities of scenarios the question arises whether the marginal distributions or the joint distributions of the experts should be aggregated. The following example shows that the results can be different:

Two experts estimate the probabilities for the success of a product development project and for the successful takeover of a rival company. Expert 1 estimates the probability for the success of the product development at 0.1 and the probability for the successful acquisition at 0.8. Expert 2 estimates the probability for the success of the product development at 0.5 and for the successful acquisition at 0.1. Both experts think that the events are statistically independent from each other. The joint probabilities can therefore be obtained by multiplying the marginal probabilities. Consequently, expert 1 estimates the case that both the product development and the acquisition are successful at the probability $0.8 \cdot 0.1 = 0.08$. Table 12-6 shows the distributions.

Table 12-6: Joint and marginal distributions

	Expert 1			Expert 2		
Development	Acquisition successful	Acquisition unsuccessful	Sum	Acquisition successful	Acquisition unsuccessful	Sum
successful	0.08	0.02	0.10	0.05	0.45	0.50
unsuccessful	0.72	0.18	0.90	0.05	0.45	0.50
Sum	0.80	0.20	1	0.10	0.90	1

The following Table 12-7 shows on the left hand side the distribution which results by averaging the joint probabilities. The marginal probabilities here result from line by line or column by column addition of the joint probabilities. On the right hand side, we find the distribution which results if the marginal probabilities estimated by the experts are averaged and the averages are then used for calculating the joint probabilities under the premise of statistical independence.

Table 12-7: Outcome of two aggregation methods

	Average of joint probabilities			Average of marginal probabilities		
Development	Acquisition successful	Acquisition unsuccessful	Sum	Acquisition successful	Acquisition unsuccessful	Sum
successful	0.065	0.235	0.3	0.135	0.165	0.3
unsuccessful	0.385	0.315	0.7	0.315	0.385	0.7
Sum	0.45	0.55	1	0.45	0.55	1

It can be seen that the joint probabilities differ significantly (admittedly, the experts' estimates in the example are so different that averaging does not really make sense). The probability that both the product development and the acquisition are successful is 6.5% in the first and 13.5% in the second instance.

For this example, there is no generally "correct" aggregation method. What may be said in favor of the aggregation of joint probabilities is that these in the end represent the relevant data. However, the assumption of statistical independence on which the experts acted is violated. Actually, the probability of a successful product development depends on the success of the acquisition: if the acquisition works out, the product development's chance of success is $0.065 / 0.45 = 0.144$, if it fails, the chance of success shrinks to $0.235 / 0.55 = 0.427$. If the two events are considered independently, this result is not very convincing. Requirements for aggregation methods for probabilities are discussed by Kleindorfer et al. (1993).

12.6.2 Aggregation of individual estimations using prediction markets

Simple averaging or averaging weighted by certain criteria can often achieve good results. Therefore, this simple form of aggregation prevails in many areas today, for instance in financial analysts' consensus estimations: the analysts' expectations of a company's earnings are averaged (mostly non-weighted) in order to arrive at a condensed overall prognosis. However, the quality of such an aggregation is often impaired by adverse factors: firstly, the interviewees might not have any incentives, or possibly even the wrong incentives for their estimates. Secondly, the individual estimations are often already strongly condensed (e.g., only the "most likely" or "expected average" of the earnings are quantified). Finally, the aggregation process does not allow for an information exchange between the experts.

So-called "prediction markets" constitute a completely different aggregation mechanism. Prediction markets are like other exchange markets where specific securities or contracts can be traded. The payoff of these securities or contracts depends on the outcome of certain future events (Wolfers and Zitzewitz 2004). Such events could be the outcome of a political election or the result of a sporting event. The contents of such a tradable contract could even be the question whether or not

anyone will press murder charges in the aftermath of Michael Jackson's death (at least in the summer of 2009, you could have traded all of the mentioned events on intrade.com). In the case of "winner-take-all"-prediction markets, the tradable contract is related to a clearly defined event (e.g. "candidate XY wins the next presidential election") or a specified interval of possible event outcomes (e.g. "DAX will close at yearend at more than 5,000 points"). These contracts promise a previously determined payoff (e.g. €1) if the underlying event occurs or the realization of the possible outcomes falls within the interval stated by the contract; otherwise, these securities become worthless. These forms of contracts are hence state-contingent claims with a price built by the trading of many market participants. Suppose the price of such a contract is €0.30. If a market participant believes the underlying event to be more likely than 30%, he could buy this contract and on average make a profit. Another market participant could sell this contract if in his opinion the likelihood for this event to occur is less than 30%. Since the traders (have to) corroborate their expectations by using their own money to buy the contracts – following the motto "put your money where your mouth is" – the resulting market prices can, under comparatively realistic assumptions, be interpreted as aggregated probabilities for the occurrence of the events generated by the market (whether or not the risk attitude of market participants needs to be considered is discussed in Wolfers and Zitzewitz 2007). For example, the price of the aforementioned Michael Jackson contract varied during the summer of 2009, depending on the information available, between 40 and 60 cents, indicating an aggregated probability estimate of 40% to 60%.

In recent years, prediction markets have become more and more important in theory and practice, not least because research results have so far attested them high predictive power which is generally higher than for standard expert interviews, polls, or surveys. Because of their structure (similar to common financial markets), prediction markets offer an efficient mechanism for aggregating dispersed information; the whole range of individual probability estimations is reflected in the price of the traded contracts. Such markets provide strong incentives for the research of information and the disclosure of their true beliefs to the market participants trading freely.

The Iowa Electronic Market (IEM) of the University of Iowa is one of the most important prediction markets. Since 1988 it has been used as a forecasting tool for the presidential elections and meanwhile offers an even broader spectrum of "election exchanges". Prediction markets have also been successfully used in corporations: at Hewlett Packard, employees implemented such markets internally to elicit probabilities for the future turnover of their printers (Chen and Plott 2002). Siemens Austria used prediction markets to estimate their order timeliness (Ortner 1997; Ortner 1998). In both cases, these predictions proved to be more accurate than the original target figures.

12.7 Making a group decision

If the individual modules of the problem have been determined by the group, a solution can be found as in the case of the decision of an individual. Instead of using the individual value, utility and probability information of a single person, the group values - resulting either from mutually shared opinions or from the mathematical aggregation of individual opinions - are taken as a basis.

In case of partial information regarding the evaluations or probabilities, dominated alternatives may possibly be eliminated, leaving one conclusively optimal alternative. If this is not the case and continuing the group process does not promise additional information, for example in terms of narrowing the intervals, the group cannot reach a collective decision.

In this case, the problem can be solved by having each member define their preferences regarding the alternatives and applying an aggregation mechanism to combine these preferences; usually, the alternatives are voted on.

Even if a group process is not concluded with a decision shared by everyone but with a decision achieved by vote, the phase of joint structuring is expected to contribute to a higher degree of rationality.

Questions and exercises

12.1

The ambassadors of two countries, Northeastia and Southwestian, are negotiating about building a facility which is meant to reduce the air pollution of both countries. Three alternatives, X, Y, and Z are shortlisted. For their evaluation, four criteria are used: 1) air improvement in Northeastia; 2) air improvement in Southwestian; 3) costs for Northeastia; and 4) costs for Southwestian. The negotiators have already agreed on the evaluation of the alternatives for each dimension (see table below). There are still different views concerning the weights. The standardized weights (meaning that the sum of weights equals 1) assigned by the ambassadors of both countries are also in the table.

(a) Which project does Northeastia prefer? Which one does Southwestian prefer?
(b) Can one project be determined to be dominant?
(c) Which project is chosen if they agree to average the weights?

	Air improvement in Northeastia	Air improvement in Southwestian	Costs for Northeastia	Costs for Southwestian
Project X	1	0.3	0.4	0.4
Project Y	0.6	1	0.4	0.7
Project Z	0	0.4	0.8	1
Weight according to Northeastia	0.2	0.2	0.3	0.3
Weight according to Southwestian	0.15	0.15	0.2	0.5

12.2

Three members of a selection committee, P_1, P_2, and P_3 evaluate three managers A, B, and C regarding their qualifications for an important job overseas. For the selection, they have agreed on three criteria X_1, X_2, and X_3. Their evaluation of the candidates and the weights for the attributes are shown in the following table.

		X_1	X_2	X_3
	P_1	0.7	0.85	0.6
A	P_2	0.75	0.8	0.7
	P_3	0.65	0.9	0.65
	P_1	0.8	0.9	0.8
B	P_2	0.6	0.7	0.8
	P_3	0.6	0.8	0.85
	P_1	0.7	0.8	0.75
C	P_2	0.5	0.7	0.95
	P_3	0.6	0.8	1.0
	P_1	0.3	0.4	0.3
Weights	P_2	0.25	0.5	0.25
	P_3	0.35	0.5	0.15

Will one of the candidates be chosen by all members of the committee?

12.3

Charles and Anna Havefight want to buy a building plot. Shortlisted are two places, Park Street and Askew Avenue. The couple disagrees. They want to solve the problem even-handedly and rationally. They agree that their objectives are a "nice neighborhood", "good accessibility", and "beauty of the site". In this context, accessibility means quick and easy access to facilities outside their house like shopping malls, cinemas etc. Charles and Anna promise to make honest evaluations. This is the result:

Charles H.	Neighborhood	Accessibility	Beauty
Park Street	0.3–0.4	0.4–0.45	0.75
Askew Avenue	0.45	0.7–0.75	0.7–0.75
Weight	1	2.5–3.5	1.2–1.8

Anna H.	Neighborhood	Accessibility	Beauty
Park Street	0.8	0.5–0.65	0.8–0.9
Askew Avenue	0.35–0.45	0.6–0.65	0.65–0.75
Weight	1	062–0.75	0.75–0.88

(a) Can Anna and Charles make a decision if each decides on their own?
(b) Is there a chance of them finding a joint solution considering the result from (a)?
(c) A few weeks later. Charles has met the neighbors from Park Street and likes them. They tell him that a new bus line from the city will be established which will allow him easy access to his workplace. Also, a fitness studio and a shopping mall will soon open nearby. This decreases the importance of the attribute "accessibility". He rethinks his evaluations and presents a new table (see below). Anna sees no reason to change her evaluations ("I told you so"). Are they now able to agree on one alternative?

Charles H.	Neighborhood	Accessibility	Beauty
Park Street	0.65–0.7	0.8	0.75
Askew Avenue	0.45	0.7–0.75	0.7–0.75
Weight	1	1.0–1.3	1.0–1.3

12.4
The sales managers Mike and Steven try to generate a joint probability distribution for the sales volume of the pop group *Cod Liver*'s first CD. First, each of them determines a distribution function by estimating upper and lower limits and the 25%, 50% and 75% points. The results are shown in the table below.

	Mike	Stefan
Lower limit	95,000	80,000
25%	115,000	105,000
Median	125,000	125,000
75%	140,000	150,000
Upper limit	175,000	200,000

(a) The aggregation should be done by using the averages of both estimates. Depict the joint density function and the joint distribution function in a graph.
(b) Alternatively, the managers decide to aggregate the individual density functions by weighting them equally. Determine the resulting joint density function and the joint distribution function.

12.5
A new toothpaste is to be tested in the federal state of Saarland before being launched nationally. Two brand experts state their estimates for the probability of the success in the testing market as: "awesome"; "okay"; and "flop". For each of these cases, they also state the conditional probabilities for the nationwide rollout being "profitable" and "unprofitable".

Expert 1	Probability	Nationwide rollout	Conditional probability
Awesome	0.2	profitable	0.7
		unprofitable	0.3
Okay	0.6	profitable	0.5
		unprofitable	0.5
Flop	0.2	profitable	0.3
		unprofitable	0.7

Expert 2	Probability	Nationwide rollout	Conditional probability
Awesome	0.3	profitable	0.8
		unprofitable	0.2
Okay	0.4	profitable	0.6
		unprofitable	0.4
Flop	0.3	profitable	0.2
		unprofitable	0.8

(a) Determine the probabilities of the six scenarios for each expert.
(b) Aggregate the experts' estimates using two different methods.
(c) What is the probability of a successful nationwide rollout considering the individual estimates as well as the aggregated estimates?

Chapter 13: Descriptive aspects of decision making

13.0 Summary

1. Descriptive preference theories try to model intuitive decision-making behavior.
2. There are many systematic deviations from behavior as predicted by utility theory.
3. We will cover four systematic deviations: incorrect probability estimations, the Ellsberg paradox, reference point effects, and the Allais paradox.
4. Descriptive preference theories provide the foundation for a wide range of economic models.
5. You will learn about preference theories that extend expected utility theory and explain many of the deviations discussed.
6. Cumulative prospect theory by Tversky and Kahneman (1992) is the most prominent descriptive preference theory. Its main attributes are reference point dependence, diminishing value sensitivity for gains and losses, loss aversion, and non-linear probability weighting.
7. Many phenomena of everyday life can be explained by prospect theory.
8. Other descriptive theories such as the disappointment theory or regret theories structurally deviate even more strongly from utility theory's evaluation principles than does prospect theory.
9. Descriptive insights into decision making also play an important role in prescriptive decision analysis. They not only help develop interview and analysis methods less prone to errors but are also relevant in interactive decision situations where rational decision making requires the correct anticipation of other players' behavior.

13.1 Descriptive preference theories and rational behavior

This book tries to provide guidance on rational decision making. The preceding chapters have presented theories and procedures that are designed to help you or a group of decision makers find the optimal alternative or arrive at a ranking of alternatives. In case you accept certain axioms as the foundation of your decision-making behavior, value or utility theory assist in deriving the right decision criteria and methods of eliciting the relevant value or utility functions.

In this chapter, we will undertake a short excursion into descriptive preference theory. As illustrated in Chapter 1, descriptive preference theory tries to capture peoples' actual intuitive decision-making behavior. The question hence arises of why we include such a chapter in a book on rational decision making:

F. Eisenführ et al., *Rational Decision Making*,
DOI 10.1007/978-3-642-02851-9_13, © Springer-Verlag Berlin Heidelberg 2010

First of all, we think that the relevance of a structured and rationality-driven approach to decision analysis can be better understood if you know more about the common mistakes of intuitive decision making. Furthermore, the descriptive insights have an immediate impact on the development and application of instruments designed for decision support. For instance, it has been become clear when determining value, utility, and probability functions that prescriptive theory is very demanding regarding the consistency of judgments of decision makers. Depending on the elicitation method applied, systematic biases could occur during the elicitation of utility functions (see Chapter 9.4.6). A certain knowledge of intuitive decision making is indispensable in order to be able to anticipate and correct behavioral problems within the framework of prescriptive decision theory.

An additional link of this chapter to the previous explanations is established by the focus on selected topics and ways of thinking within descriptive decision theory. We neither can nor want to tackle the entire range of descriptive theories. From its early stages on, much work in psychology has dealt with explaining and predicting human decision making. This chapter will stick closely to the approach of the previous chapters. We will reflect on systems of axioms time and again, and derive preference theories from these systems. In contrast to value and utility theory, the axioms and theories presented in this chapter are not the cornerstones of rational behavior but try to capture intuitive decision-making behavior; that is why we talk of descriptive preference theories. The approach of reflecting on axiomatic frameworks in the domain of descriptive theories as well is geared towards our goal of decision support. It allows for a fundamental understanding of the differences between prescriptive and descriptive decision-making behavior (what axioms are modified and how?). This part of decision theory is also referred to as *behavioral decision theory*.

The reasons we have presented so far, of course do not have to convince you to read this chapter. If you are only interested in learning about rational decision making, you should know something about systematic biases but descriptive preference theories are irrelevant to you. The following information is intended to familiarize you with some of the most interesting developments in descriptive decision theory. We hope you share the experience of many researchers in the field of economics: over the last years, they have developed a deeper interest in descriptive preference theory as they have realized that being able to model actual decision-making behavior mathematically is beneficial to the derivation of economic theories. Some simplified examples are intended to prove this assertion.

Marketing tries to predict the market share of products. In order to do so, it is helpful to have theories available to predict consumer choice behavior. Stock markets are another example; the market is nothing but an aggregation of supply and demand for specific types of securities. If the buying and selling behavior of investors is predictable with the help of descriptive decision research, economic models could be developed that better describe the formation of prices for securities and perhaps also explain the existence of price bubbles. Agency theory is a central element of modern organizational theory. One of the main goals of this strand of research is to arrange the behavior of the agent by the use of incentive systems in such a way that the agent maximizes the utility of the principal. Whether incentive systems based on utility calculus (to model the behavior of the agent) really accomplish the intended effect is a question to be answered within the realm of descriptive decision research.

The list of areas suitable for applying descriptive preference theories could easily be extended. However, this chapter does not only showcase applications of descriptive theory but places its emphasis on presenting some descriptive theories themselves (for an application of descriptive theories in investment and financing theory see, e.g. Weber 1990; Weber 1991; for an application in banking theory see, e.g. Langer 1999). Hence in the next section of this chapter we will first present typical intuitive decision behavior that deviates from rational behavior as predicted by utility and value theory. We will then introduce descriptive decision theories which claim to explain at least some aspects of the deviant behavior. In particular, we will focus on the cumulative prospect theory by Tversky and Kahneman (1992) since it is currently the most prominent descriptive decision theory under uncertainty.

13.2 Examples of intuitive behavior not in line with utility theory

This section discusses in detail four effects that nicely illustrate the discrepancies between intuitive preferences and preferences as modeled in utility theory. When selecting these effects, two considerations have guided our way. On the one hand, we want to demonstrate that these problems can occur at various stages of the decision process. For illustrative purposes, Figure 13-1 graphically reviews the stages of a decision process according to subjective expected utility theory. The circles numbered 1 to 4 point out the stages where the problems to be discussed in the following can occur; we will also present them in this order. On the other hand, the selected phenomena are characterized by accompanying (descriptive) theories designed to model their occurrence (see Section 13.3). Some of these phenomena – often referred to as paradoxes – have already been discussed at an earlier point in this book. Following the discussion on the four selected effects, this section closes with a brief overview of further phenomena that are not in line with utility theory.

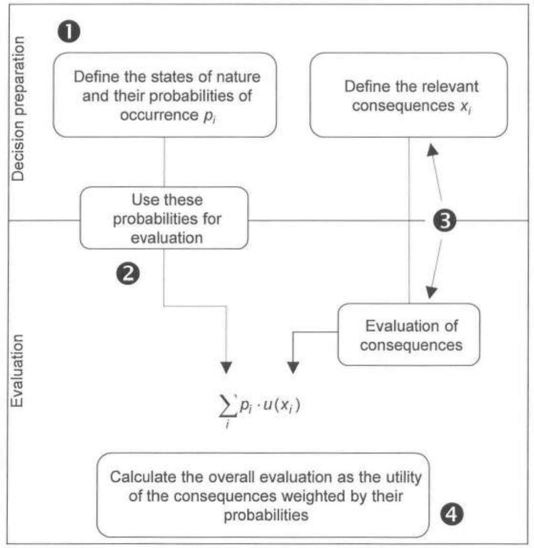


Figure 13-1: Sequence of events in the SEU decision process and matching of phenomena

13.2.1 Bias in probability estimates

An important step during decision preparation is to determine the relevant states of the world and their probabilities of occurrence. However, as already discussed in Chapter 7, decision makers have problems with consistently judging probabilities. It is hence not surprising that a plethora of paradoxes when dealing with probabilities have been documented to date. We will now present a class of problems that led to the development of support theory as presented in Section 13.3.1. Several additional paradoxes from this area are surveyed in Section 13.2.5.

It sounds trivial that the subset of a set of events cannot have a higher probability than the event set itself. A violation of this property is convincingly documented in the following examples (Tversky and Kahneman 1983):

You have a six-sided die that has four red and two green sides. The die is rolled twenty times and the sequence of red (r) and green (g) is recorded. In the following we provide you with three sequences a, b, and c. You receive €50 if the sequence you choose appears within the total sequence of die rolls. What sequence do you choose?

a: r g r r r
b: g r g r r r
c: g r r r r r

Most decision makers choose sequence *b* (63%) and only 35% choose sequence *a* (the remaining 2% chose *c*). You – by now skilled in probability theory – immediately recognize that the probability of occurrence of sequence *a* must be greater than the one of sequence *b*. To obtain sequence *b*, sequence *a* and additionally the (prior) event "die shows green" must occur. Yet, as the probabilities of 2/3 for red and 1/3 for green are exactly reflected in sequence *b*, most decision makers choose the disadvantageous sequence *b*. Sequence *b* is especially *representative* of the generating random process and is hence judged to be especially likely.

In addition, the degree of detail in the description of an event should not have an influence on the probability estimate. Fischhoff et al. (1978) asked experts and laymen to quote the probabilities for different reasons for a car to fail to start. The residual "other problems than battery, engine or fuel system" was given a probability of $p = 0.22$. When the residual was characterized more extensively as "other problems than battery, engine or fuel system, e.g. starter or ignition," the assigned probability increased to $p = 0.44$. The extended details of the description of the event led to an increase in the probability estimate.

A study by Fox and Clemen (2005) shows that even experienced decision analysts are not immune to their probability estimates being inconsistent. They asked members of the Decision Analysis Society (DAS) to assess the probability that the total number of members of their society would be more or less than 1000 five years in the future (the current number of 764 members at the time the study was conducted was known to everyone). Half of the participants were asked to assign likelihoods to the events of membership falling into each of the intervals [0, 400], [401, 600], [601, 800], [801, 1,000], [1,001+]; the other half were asked for likelihoods for the membership intervals [0, 1,000], [1,001, 1,200], [1,201, 1,400], [1,401, 1,600], [1,601+]. To rule out the possibility of the participants drawing any conclusions from the arbitrary definition of intervals and presuming hidden information ("obviously the well-informed organizers of this study expect membership numbers to rise dramatically, otherwise they would not present all these intervals exceeding 1,000"), both groups were informed of the differently cut intervals presented to the other group. Nevertheless, the judged probability that the DAS would have more than 1,000 members differed highly across groups. While the first group (with only one interval exceeding 1,000) stated a probability of 10% on average, adding up the four probability judgments of the second group produced an overall probability of 35%.

13.2.2 The Ellsberg paradox

Subjective expected utility theory does not only require a rational decision maker to assign probabilities to the relevant states of the world but also that the origin of the probability judgments does not play a role in the subsequent evaluation procedure at all. The probability estimate "40% probability" made by an expert has to

be treated identically to the 40% estimate of laymen. In actual decision-making behavior it can be observed however that people evaluate alternative sources of uncertainty differently (source preference). Ellsberg (1961) nicely illustrates this point with the Ellsberg paradox named after him. This paradox is based upon an urn that contains 30 red balls and *a sum* of 60 black and yellow balls, without the proportion of black to yellow balls being known. Any proportion of black to yellow balls is possible; they only have to add up to 60 balls in total. Again, two alternative pairs of lotteries are to be evaluated with all the potential gains being identical.

a: You win if a *red* ball is drawn from the urn.
b: You win if a *black* ball is drawn from the urn.

and

a': You win if a *red or yellow* ball is drawn from the urn.
b': You win if a *black or yellow* ball is drawn from the urn.

Most decision makers prefer *a* to *b* and *b'* to *a'*. This is often justified by pointing out that for *a* and *b'* the probability of winning is known with certainty, being 1/3 and 2/3, respectively. Yet, something like a "different certainty" of probabilities is unknown to subjective expected utility theory. Instead, subjective expected utility theory clearly demands a decision maker to prefer either *a* and *a'* or *b* and *b'* or to be indifferent in both cases. The commonly observed preference $a \succ b$ and $b' \succ a'$ directly violates the independence axiom of the subjective expected utility theory (*sure thing principle*). The lottery pair a, b differs from the lottery pair *a'*, *b'* just by also including winning for the event "a yellow ball is drawn." As the axiom requires this event to be irrelevant to the preference, the preference must be identical for both lottery pairs.

Several alternative explanations exist for the behavior as demonstrated by the Ellsberg paradox. Not knowing the probabilities could be regarded as an additional risk factor or the level of reliability of a probability could serve as an explanatory factor. Many decision makers intuitively desire to assign different levels of reliability to subjective probabilities. The event "red ball is drawn" carries a subjective probability of 1/3 with a high level of reliability. If a subjective probability for the event "red or yellow ball is drawn" is quoted, many decision makers feel uncomfortable: one wants to assign to this subjective probability a low level of reliability. The aversion to choosing an alternative where one is unsure about the probability is also referred to as ambiguity aversion.

The importance of the Ellsberg paradox extends far beyond the urn example that was just discussed. In Chapter 7 we have seen that virtually all cases of decision making under risk require subjective probabilities as a basis. The decision maker will be more confident about some subjective probabilities than others, but this must not have any consequences for decision-making behavior according to SEU theory. For example, think of the decision problem of opening up a wine pub in the "golden Pfalz" (an area in southwest Germany) or in Inner Mongolia. You would have estimated the probability of success or failure to be $p_{success} = p_{failure} = 0.5$ for either alternative. However, at least one of the authors of this book has a bad feeling about the probabilities regarding Inner Mongolia: he attributes a lower level of reliability to these probabilities than to the probabilities for the "golden Pfalz." An aversion to lower levels of reliability – in the spirit of the Ellsberg paradox – may well influence decision-making behavior, but this is not in accordance with SEU theory. The example also illustrates that ambiguity aversion is subjective and *not* a property of the event that is to be evaluated. A reader from Inner Mongolia would probably attribute a lower level of reliability to the probabilities for opening the pub in the "golden Pfalz."

13.2.3 Reference point effects

Another important stage in the decision process is the definition and evaluation of the consequences. Throughout this textbook, we have talked extensively about consequences while spending little time on the different ways of their presentation and the resulting possibility of differences in perception and evaluation. Think of

- the famous glass of water that can be viewed as "half full" or "half empty" depending on one's perspective;
- a pay raise of 3% that can seem generous (you did not expect anything) or disappointing (your colleague just received 6%); or
- your 2-room apartment that seems luxurious when compared to your student dorm, but then again humble when compared to your friend's villa.

All of these examples have in common that the consequences are evaluated with respect to a reference point and viewed as gains or losses with respect to this reference point. The formal "discovery" of the concept of the reference point is maybe the most important development of descriptive decision theory. You must realize that this issue raises the fundamental question of which components of a decision problem actually constitute the relevant input for the evaluation procedure. The traditional way of thinking as reflected by utility theories demands that final wealth levels are to be evaluated. A decision maker thinks, for instance, of the final wealth positions that can be generated from various alternatives and of the utility they create. Yet, from a descriptive point of view, it seems that decision makers instead evaluate changes (or rather differences with respect to a reference point). The decision maker thinks for instance of the changes relative to the current wealth (or any other reference point) that an alternative would bring about. It goes without saying that the location of a reference point can have an impact on decision-making behavior. We will refer to this impact as the *reference point effect*. For example, decision makers generally hate losses much more than they like gains; this is why we also talk of *loss aversion*. It can also be documented that the risk attitude of a decision maker is influenced by whether he thinks about gains or losses.

We will demonstrate the importance of the phenomenon with the help of an example where different locations of the reference point lead to different decisions in situations that are identical in economic terms (analogous to Kahneman and Tversky 1979). Compare the following situations:

Situation 1
You possess €500 and face the following alternatives:

$$a=(€0, 0.5; €100, 0.5) \quad \text{vs.} \quad b=(€50, 1)$$

Situation 2
You possess €600 and face the following alternatives:

$$a'=(-€100, 0.5; €0, 0.5) \quad \text{vs.} \quad b'=(-€50, 1)$$

Many decision makers will choose the sure payment b in situation 1, whereas they will choose the lottery a' in situation 2. As the expected payoffs are identical, they exhibit risk aversion in situation 1 and act as risk loving in situation 2. Looking at both situations in terms of wealth levels, you will notice that they are identical. In each case, you have to decide upon a risky alternative where you have a 50/50 chance of a wealth level of €500 or €600 and a sure wealth level of €550. Because of the decision makers' tendency to evaluate wealth changes rather than the resulting (final) wealth levels, the different locations of the reference point (which is located at the status quo of €500 and €600, respectively) lead to differences in behavior across both scenarios.

Even if the initial situation and the differences to choose from are absolutely identical, the presentation of alternatives can create reference point effects. To demonstrate this, we will slightly modify the above situation 1:

Situation 3:
You possess €500 and face the following alternatives:
You first receive a sure payment of €100 and can then decide between

$$a'=(-€100, 0.5; €0, 0.5) \quad \text{vs.} \quad b'=(-€50, 1)$$

Now, situations 1 and 3 do not even differ in the initial wealth. The only difference between the scenarios is that in situation 3, the alternatives are split into two components (a sure payment and a choice among loss lotteries) whereas in situation 1 the payments are aggregated into one gain lottery. A careful decision maker will immediately recognize this equivalence and mentally transform situation 3 into situation 1. However, many decision makers intuitively act risk loving in situation 3 and risk averse in situation 1. The sure payment of €100 is not integrated into the lottery but into the reference point ("I now own €600 and must decide between two loss lotteries"). It is a classic example of a framing effect: the presentation of a decision problem has an influence on the decision.

Let us look at a second example of reference point effects and address the question of the maximum price a decision maker is willing to pay for a product (buying price) and the minimum price at which he is just willing to sell the same product (selling price). Essentially, both cases raise the question of what utility the ownership of the product conveys to the decision maker and ask for the change in monetary wealth creating comparable utility. Both prices should thus be identical for sure goods (goods with sure outcomes), and can differ only moderately for risky goods (goods with risky outcomes, e.g. stocks). However, when decision makers are asked for their respective buying and selling prices, surprising differences are revealed: the demanded selling price is generally much higher than the buying price one is willing to pay. An overview by Kahneman et al. (1990) shows that the ratio of selling price to buying price strongly varies depending on the type of question asked but is larger than two in almost all studies. You can obtain an idea of this discrepancy in prices when trying to come up with your own prices for the following two situations:

- How much are you willing to pay at most for avoiding the chance of suffering personally from a severe illness with $p = 0.1\%$?
- How much must one at least pay you for you to accept the risk of personally suffering from a severe illness with $p = 0.1\%$?

Typically, respondents demand more than the tenfold amount to take an additional medical risk compared with the situation in which they can reduce the risk by an equivalent percentage amount (Thaler 1980). However, meaningful discrepancies between buying and selling price can also occur for sure goods. Put yourself into the situation of a tennis fan who receives tickets for the Wimbledon final as a present. For what amount would you be willing to sell the tickets? Compare this selling price with the price you would be willing to pay for the tickets.

Thaler (1980) has labeled the divergence of buying and selling price as the *endowment effect.* This effect could also be characterized by the sentence: "Once I own something, I don't like to give it away!" Decision makers judge goods not in absolute terms but relative to a reference point. When making a purchase, you receive the product but give up wealth which leads to a smaller buying price due to the burdening of loss aversion. When selling, you lose the product and this burden can again only be compensated by a high monetary reward (high selling price). The reference point concept is unknown to utility theory and thus cannot explain differences in behavior for gains and losses with respect to a reference point (see Weber 1993 for an overview of endowment effects).

13.2.4 The Allais paradox and certainty effects

The independence axiom is one of the core components of expected utility theory. It provides for the additive form of the value function where probabilities have linear impact. As has been illustrated in Section 9.2.3 with the help of the probability triangle, this axiom substantially reduces the set of preferences conforming to utility theory. It would thus be surprising if decision makers were not to violate this axiom intuitively. Allais (1953) defined examples of lottery choices in which most decision makers will violate the independence axiom. Figure 13-2 presents such an example that has already been reviewed in Section 9.2.3 under the label of the Allais paradox. A decision maker is confronted with two pairs of lotteries for which the independence axiom mandates the preference regarding the first pair (a and b) to match the preference regarding the second pair (a' and b'). The second pair is derived from the first pair by mixing the latter with the identical and thus irrelevant (according to the independence axiom) third lottery ($c = (€0, 1)$). It holds that: $a' = 0.25\ a + 0.75\ c$ and $b' = 0.25\ b + 0.75\ c$.

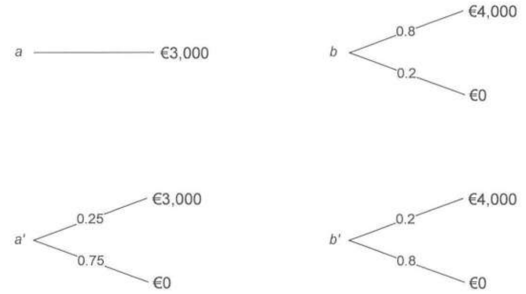


Figure 13-2: The decision situation in the Allais paradox

Most decision makers who must decide between the alternatives as presented in Figure 13-2 prefer a to b. They like the sure amount of €3,000 more than the risky chance of winning €4,000. For the second pair, the majority of decision makers prefer b' to a' because, given a choice between two risky prospects, one prefers the alternative with the higher gain (higher expected value). As already mentioned, this behavior is not compatible with utility theory but is nevertheless widely observed in intuitive decision making.

For the time being, we want to set aside the formal reasoning (caused by the independence axiom) and try to grasp the intuition behind this violation of rational behavior. The ratio of the probability of winning to the probability of not winning is equal to 5:4 for both the paired comparison of a vs. b (100% to 80%) and for the paired comparison of a' vs. b' (25% to 20%). the relative attractiveness of the lotteries thus should not change. The decision maker, however, perceives the difference between *sure* (100%) and *pretty sure* (80%) to be greater than the difference between *not really probable* (25%) and *even a bit less probable* (20%). Such a bias in the perception of probabilities (or probability weighting) is not allowed for in expected utility theory where the probabilities have to enter linearly into valuation. In particular, there cannot be a certainty effect of the kind "the transition from 99% to 100% is especially important to me, much more important than expressed by a 1% difference in probabilities."

This violation of the independence axiom caused Allais (1953) to dismiss the "American school" of preference theories based on von Neumann and Morgenstern and Savage. Allais (1979) introduced his own prescriptive theory, but it never became widely accepted: most decision makers do not want to violate the independence axiom in rational decision making once they have given it some thought.

In the context of descriptive preference theories, the Allais paradox's systematic deviating from utility theory needs to be accounted for. A variety of approaches can explain the preferences underlying the Allais paradox. For some of these approaches, fully fledged theories that are capable of describing the "paradox" decision-making behavior have been developed. In a probability triangle, these theories must generate indifference curves that differ from the straight parallel lines of expected utility theory. Reconsider Chapter 9, especially the insights from Figure 9-9. At that point, we had already realized that parallel indifference lines (independent of their slope) are unable to explain the typical decision pattern of the Allais paradox, namely $a \succ b$ and $b' \succ a'$. Figure 13-3 illustrates what alternative indifference curves that have this kind of explanatory power could look like. As you can see, it is not mandatory to give up the property that the indifference curves are straight lines. This problematic preference pattern can be explained by straight lines that *fan out* from the origin (x_m). Hence, it is also not mandatory to sacrifice the betweenness property that goes along with linear indifference curves. *Betweenness* holds if from $a \sim b$ follows $a \sim p \cdot a + (1-p) \cdot b \sim b$, where a and b are lotteries and p is a probability; in other words: if a decision maker is indifferent between the two alternatives a and b he also must be indifferent between a, b, and any convex combination of a and b (that lies "in between" a and b).

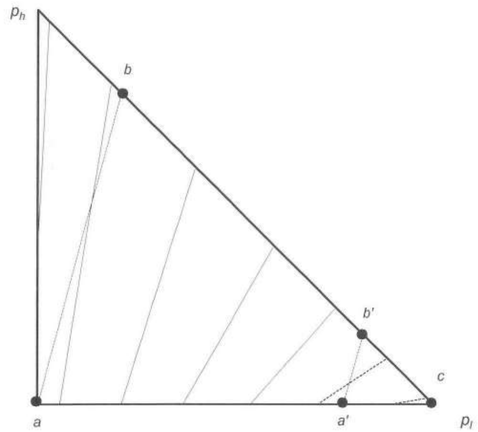


Figure 13-3: Three-outcome-diagram for the Allais paradox

13.2.5 Overview of decision behavior phenomena

In the following, we will give an overview of further aspects of human decision making that are highly relevant to the descriptive decision theory but generally are not grouped under the umbrella term "rational behavior." We have tried to define each effect briefly, state its source of origin, and cite – if possible – a recent article summarizing the relevant research literature dealing with the effect. Many effects were and are mentioned in other sections of this book and partly discussed in detail.

However, for reasons of completeness we will once again list them in this section. The effects are presented in alphabetical order.

Attitude towards ambiguity

When evaluating an uncertain alternative, the decision maker can either be sure or unsure about the subjective probabilities with which the individual consequences are believed to be occurring. There is, for instance, little uncertainty about the fact that the probability of "heads" is 50% when flipping a coin. When drawing a ball from an unknown urn containing red and black balls, one must also assume a probability of 50% for "red" if no additional information is available. Decision makers are nevertheless unsure of this probability assessment; we speak of a lottery that is afflicted with ambiguity. Individuals in general are ambiguity averse (Ellsberg 1961), i.e. they find more ambiguous lotteries less attractive, even if the probabilities and consequences of the lotteries are identical. For an overview of the current state of research, we refer to Camerer and Weber (1992); see also Sections 13.2.2 and 13.3.3.1.

Anchoring and adjustment

People who need to make decisions under uncertainty make use of this heuristic coined by Tversky and Kahneman (1974) if their initial assessment of an uncertain decision variable is based on a starting point (anchor) and then gradually adjusted (adjustment). However, the anchor can be determined by an external standard preset that does not contain any information on the true value of the decision variable. Slovic and Lichtenstein (1971) show that during the assessment of unknown variables, people insufficiently adjust from the starting point towards the true value. See also Epley et al. (2004) and Section 7.5.4.

Availability bias

The availability bias or availability heuristic (Tversky and Kahneman 1973) is a judgment bias caused by a memory psychological effect. It is the result of the fact that the easier or faster it is to recall examples of an event from memory (high availability), the higher the subjective probability that people assign to its occurrence will be. However, if the availability of an event has nothing to do with its likelihood but instead is related to its special emotionality, high familiarity, high media coverage, or its occurrence being recent, this heuristic leads to an overestimation of an event's probability, causing biased decisions. See also Section 7.5.2.

Base-rate fallacy

A formally correct consideration of newly arrived information while processing probabilities, i.e. the transition from a priori to a posteriori probabilities, requires the application of Bayes' Theorem. If people deal with probabilities in a more informal and intuitive way, they tend to give too little weight to the base rate (i.e. the a priori probability) as compared to the newly arrived information (Kahneman and Tversky 1973). Camerer (1995) gives an overview of some experimental studies related to this phenomenon. See also Section 7.5.3.

Certainty effect

The certainty effect describes the phenomenon that decision makers consider the difference between two probabilities in their decision as especially strong if it is a transition from "almost certain" to "certain." For example, an increase in the probability of winning a lottery by 1% is especially desirable if the winning probability is thereby no longer 99% but 100%. In contrast, raising the probability from 30% to 31% is perceived as much less important. This phenomenon was used by Allais (1953) to construct lotteries where decision makers intuitively violate the independence axiom of expected utility theory (for the Allais paradox see Sections 13.2.4 and 13.3.2).

Choice vs. matching anomaly (response mode bias)

The *response mode* refers to the procedure a study uses to elicit decision makers' preferences. These in particular include *choice* tasks, where a decision maker for instance chooses between a set of lotteries, and *matching* tasks, where preferences are expressed by providing a certainty equivalent for a given lottery, for example. In contrast to the postulate of (procedural) invariance Tversky et al. (1988) as well as for instance Ahlbrecht and Weber (1997) show that the observed preference can depend on the mode of elicitation (response).

Disappointment effect

Whether winning €100 in a risky 50/50 lottery is perceived as an elating or disappointing outcome depends on whether the alternative was winning €0 or winning €200.

Anticipating the disappointment of missing a hoped-for gain can cause decision makers to discard choosing an alternative with a chance of a high gain ex ante (Bell 1985). Thus, the evaluation of the individual consequences of an alternative is not carried out independently but with respect to the other potential consequences (lottery-dependent utility; Loomes and Sugden 1986). See also Section 13.3.3.

Endowment effect

The endowment effect describes the phenomenon that the valuation of an object depends on whether or not people own the object. The minimum price a seller is willing to accept for a specific object is located above the maximum price he would be willing to pay as a buyer. See Thaler (1980), Weber (1993), Sayman and Onculer (2005), and Section 13.2.3.

Framing effect

The frame of a decision situation refers to the design of the aligned alternatives, states of the world, and outcomes. Framing effects occur because people arrive at different decisions for the same problem depending on how the problem is framed. The most famous example is given by Tversky and Kahneman (1981) and is related to framing the outcomes as gains and losses causing the decision makers to react differently.

Gambler's fallacy

When asking roulette players for the color they expect to show after seeing "red" ten times in a row, they typically answer "black." This behavior, the opinion of expecting the other color to show after a series of one color, is called *gambler's fallacy*. It can be attributed to the representativeness heuristic as many people believe that such a long series of only one color would not be representative of the underlying random process. See Tversky and Kahneman (1971) and Croson and Sundali (2005).

Hindsight bias

According to Fischhoff (1975), this bias rests on two facts. On the one hand, the announcement of the occurrence of a specific event retrospectively increases its perceived ex-ante probability of occurrence. On the other hand, people who have received information on the occurrence of the event are not conscious of this fact; the change in their perception is carried out subconsciously. As a result, individuals overestimate the degree of agreement of their judgments before and after the occurrence of an event. They think they knew all along what would happen (the *knew-it-all-along* effect or the curse of knowledge). See also Guilbault et al. (2004) and Section 7.5.2.

Illusion of control

People often believe they have more control over random events than actually is the case. This effect, referred to as "illusion of control" (Langer 1975) is especially pronounced in situations that people regard as familiar or where they can actively participate, like for instance choosing between alternative lotteries. It leads to the over- or underestimation of probabilities of events' occurrence as well as to a preference for the alternative that offers more control.

Loss aversion

Kahneman and Tversky (1979) show in their studies that people suffer more from losses than they feel happy about gains of an equal size. Therefore, the value function assumed in prospect theory shows a steeper slope in the area of losses than in the area of gains (relative to the reference point). Loss aversion hence also implies that decisions depend on the framing of the alternatives' outcomes, e.g. as gains or losses (see Camerer 2005) and Sections 13.3.3. and 13.3.3.4.).

Mental accounting

Individuals are prone to "mental accounting", treating different wealth positions as if they were in different accounts (Thaler 1985, Thaler 1999). For instance, many people divide their securities deposit into the mental accounts "retirement" and "speculative investments." Mental accounting refers to the fact that decisions are optimized only within the corresponding mental account while possible interdependencies with positions in other accounts are ignored. In doing so, decision makers violate the principle of "asset integration" demanding that a decision should be made considering the overall wealth level. Furthermore, the myopic evaluation of long-term investments (narrow bracketing) could also be seen as a form of mental accounting (for myopic loss aversion, see Section 13.3.3.5).

Omission bias
The so-called omission bias describes the subjective perception that acting per se is considered to be riskier than doing nothing (omission). This can for instance, explain why parents are often reluctant to have their children vaccinated although the probability of an illness is demonstrably reduced by the vaccination. Closely related to the omission bias is the "default bias" that captures people's propensity to choose the default option if provided. See Asch et al. (1994) and Baron and Ritov (1994).

Overconfidence bias
A series of empirical studies has shown that people tend to overestimate their knowledge and skills. For instance, 82% of the students interviewed believed that they belonged to the best 30% of drivers (Svenson 1981). This so-called overconfidence bias can also help explain the high trading volume on financial markets. It is not only found among laymen (students) but also – at least as strongly pronounced – among professional players on the financial markets (Glaser et al. 2007). Glaser and Weber (2010) give a current overview (see also Section 7.5.4).

Overestimation of small probabilities
It has been found that small probabilities tend to be overestimated when subjectively perceiving probabilities. Prospect theory accounts for this perception bias by a type of probability weighting function that in particular puts too much weight on small probabilities for extreme outcomes. See the detailed discussion in Section 13.3.3.

Partition dependence)
An individually estimated probability distribution of a continuous variable such as for instance a stock's closing price can be influenced by the specific bracketing of the continuous variable into discrete intervals; presenting an individual with two separate sub-intervals $[I_1, I_1+x)$ and $[I_1+x, I_2]$ instead of one larger interval $[I_1, I_2]$ increases the aggregate subjective probabilities thus allocated to the overall interval. Partition dependence could also be seen in the case where resources instead of probabilities are to be assigned to different categories. See Fox et al. (2005), Bardolet et al. (2009), as well as Sections 7.5.4 and 13.2.1.

Preference reversal
Depending on the mode of elicitation, preferences can change in a way that one finds a preference for X over Y using procedure a and a preference for Y over X using procedure b (the reversal effect). Lichtenstein and Slovic (1971) have shown that decision makers when choosing between lotteries quoted a lower selling price for the selected lottery than for the lottery they did not choose. See also Grether and Plott (1979).

Range effect
Within a multiattribute decision problem, the range of outcomes plays an important role when assigning objective weights (scaling constants). Changes in the range of outcomes should result in clearly prescribed adjustments of the objective

weights. However, there is empirical evidence based on experiments (von Nitzsch and Weber 1991) that decision makers show little sensitivity to the size of outcome intervals, i.e. they hardly respond to changes in the range of outcomes when determining objective weights, or don't react at all. See also Section 6.7.1.

Reference point effect
An important phenomenon in descriptive decision theory is the observation that decision makers evaluate the outcomes of the disposable alternatives relative to an individual reference point that is set in advance. For instance, investors tend to evaluate the current price of their stocks relative to their buying price or their highest historic price (Odean 1998). The consideration of a reference point is an important element in the prospect theory developed by Kahneman and Tversky (1979). See also Section 13.2.4.

Regret effects
People often judge the quality of their decisions by what would have happened if they had decided differently. According to Loomes and Sugden (1982), regret theory rests on two fundamental assumptions. On the one hand, people experience feelings such as regret or joy. On the other hand, people anticipate these feelings when they have to make decisions under uncertainty. Consequently, in addition to the utility of the consequences, people account for the emotions that result from comparing the realized consequence with the one they missed out on and they try to avoid regret; they neglect the fact that adverse consequences do not necessarily result from bad decisions. See also Section 13.3.4.

Representativeness heuristic
People often tend to be use representative characteristics of the population as an orientation point when estimating the probability of an event. A similarly important role is played by the typical characteristics of a subset of the population when deciding upon the affiliation of a sample to this subset. For instance, a series of heads, tails, heads, tails seems to represent a fourfold coin throw better than the series tails, tails, tails, tails does. This heuristic can cause biases and therefore wrong decisions if the characteristics do not reflect the actual probabilities given a priori. See Kahneman and Tversky (1972) Tversky and Kahneman (1983) Tversky and Kahneman (1983), Tversky and Kahneman (1983), and Section 7.5.3.

Splitting bias
If a system of objectives is refined by splitting a superior objective into several subordinate objectives, the sum of weights of the subordinate objectives should equal (when appropriate ranges are chosen) the initial weight of the superior objective. In reality however, decision makers generally allocate a larger sum of weights to the branch in the system of objectives that was split (Weber et al. 1988). Similar effects can be observed when splitting causal trees and for any trees displaying different states of nature (*event-splitting*; Humphrey 1996). See also the entry on partition dependence and Section 6.7.2.

Status quo bias

The status quo bias initially documented by Samuelson and Zeckhauser (1988) refers to the fact that the distinction between alternatives that change or maintain the status quo can affect individual's decision behavior. In prospect theory, this effect can be interpreted as the reference point effect or as the endowment effect.

Sunk costs

Sunk costs arise when past temporal, financial, or other types of investments lead people to make decisions they would not have made otherwise. Decisions on the continuation of projects are especially affected by this bias: "The fact that no major dam in the United States has been left unfinished once begun shows how far a little concrete can go in defining a problem" (Fischhoff et al. 1981, p. 13).

Winner's curse

The winner's curse is the result of a judgment bias in an auction context where individuals bid for an object whose value is the same to everyone but unknown. Every participant must form his own opinion on this value in the process. The higher the personal valuation, the higher the bid; the most optimistic appraiser (whose bid generally exceeds the value of the object) thus tends to win the auction. According to Thaler (1992), the winner's curse results from the systematic failure to include this problem of adverse selection in the amount of the own bid.

13.2.6 Importance of preferences deviating systematically from utility theory

Within the descriptive approach, it is necessary to distinguish whether the decision process should be described or only the decision outcome should be predicted. When mapping the decision process, one has to consider how decision makers arrive at evaluations, how these are linked, and how the choice of alternative is made. For this, insights from research in psychology play an important role. Considering the decision process, it does not necessarily have to be assumed that decision makers have exact preferences in simple decision situations. In fact, preferences can be constructed during the decision-making process (this is the common point of view in psychology, see Weber and Johnson 2009) and therefore strongly depend on the relevant circumstances when the decision maker is interviewed (point in time, elicitation method, framing of the problem, etc.). The decision maker applies certain heuristics (availability heuristic, representative heuristic, anchoring, and adjustment) in order to make value judgments or estimate relative frequencies or probabilities. For many economic applications it is however sufficient for a descriptive theory to predict the decision outcome even if the decision process might not proceed as modeled in theory. A good example here is the probability weighting incorporated into cumulative prospect theory which you will become more acquainted with in Section 13.3.4.1. It should not be assumed that decision makers actually calculate decision weights according to the formulas (13.7) and (13.8). However, these formulas adequately describe the decision patterns that arise in the case of intuitive decision making under uncertainty. In a descriptive theory, it is necessary to consider a reference point. Furthermore, violations of the

substitution axiom as well as systematic problems in determining subjective probabilities need to be reflected.

For prescriptive decision theories, it is rather comforting that intuitive behavior deviates from rational behavior. For instance, Savage was confronted – supposedly at lunch – by Allais with the paradox that today is named after him and further decision situations. At first, Savage behaved irrationally in terms of his own theory, amusing Allais who rejected the so-called American school. After careful consideration, Savage recognized the irrationality of his behavior and was glad that utility theory (especially the independence axiom) prevented him from acting irrationally. Savage even suggests systematically examining decision situations where intuitive preferences are not in line with the temporarily accepted normative theory in order to decide clearly whether the intuitive preferences need to be dismissed or maybe the axioms of the normative theory need to be adjusted (Savage 1954, p. 102). The opinion – also represented by this book – that intuitive behavior deviating from rational behavior is not at all disturbing but conveys many interesting insights will be illustrated once again by an optical illusion as an example. Have a look at the two lines in Figure 13-4 and guess which of these is longer. It has been shown in many repetitions of this experiment that the upper line is believed to be longer even though both lines are of equal length. However, it is not true that this optical illusion makes redundant the concepts of measuring the length of objects or of distance; the opposite is the case. The problem of individually biased perceptions demands a rational procedure. Measuring length with the help of a folding or metering rule are examples of such a rational procedure. It is exactly the same in decision-making research where a prescriptive theory is needed to help us arrive at an optimal decision fulfilling the standards of carefully defined rationality while taking the well-documented behavioral biases into account.

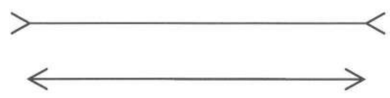


Figure 13-4: Optical illusion

The question of the extent to which the occurring irrationalities should be considered in modeling within the science of business administration or economics has led to meaningful controversies, especially in the 1980s and 1990s. Even today, scientists are sometimes still attributed to the *behavioral camp* or to the *rational camp* whereas the clear boundaries between the two camps are diminishing more and more over time. A special issue of the Journal of Business (Hogarth and Reder 1986) gives an overview of the discussion results at that time. However, an article by Thaler (1986) not only summarizes the most important pros and cons of behavioral insights that are relevant to economic modeling but also nicely reveals the entrenched positions. Thaler here suggests (hardly being serious) that using a numbered list of such pros and cons could save a lot of time at conferences because everybody could just refer to the numbers of the arguments (which are repeated over and over again anyway) instead of presenting them completely.

Advocates of utility theory, of course, know about the systematic deviations. However, they believe that these deviations are irrelevant when it comes to real economic situations (Smith 1991). Among other things, it is argued that decision makers are disciplined towards rational behavior by markets, that results from laboratory experiments cannot be transferred to reality without further ado, that sufficiently high monetary incentives would lead decision makers back to utility theory or simply that the irrationality of some market players has no impact on truly relevant economic variables (the latter argument is especially relevant to the field of financial research with its markets being regarded as efficient; see Barberis and Thaler 2003).

All the aforementioned arguments certainly make some reasonable points but are not comprehensive enough to be disproved by carefully conducted experiments (Langer 2007). The incentive argument has caused researchers to conduct their experiments in developing countries, enabling them to offer high monetary incentives from the participants' point of view (Kachelmeier and Shehata 1992). Camerer and Hogarth (1999) run a meta-analysis examining a multitude of experimental studies and conclude that on average, higher monetary incentives do not improve the decision quality. Another problem with the points mentioned above is the external validity (i.e. the question of whether or not phenomena proven in artificial experiments can be transferred to decision making in reality), which has lately been increasingly considered by conducting so-called field experiments. In these experiments, decision patterns are examined that are not perceived as part of an artificial study by the experiment participants (Harrison and List 2004). In addition, the question of whether more experience and skills (i.e. learning effects) actually reduce the typical mistakes and biases can be examined by comparative studies considering students and experts and has been refuted (Haigh and List 2005, Glaser et al. 2010). Finally, the assumption that the irrational behavior of individual market players is not reflected in market data has been refuted both theoretically (Russell and Thaler 1985, Shleifer and Vishny 1997) as well as by experimental studies It has been shown that the Ellsberg paradox (see Section 13.2.2) is not only reflected by market prices (Sarin and Weber 1993) but also causes systematic changes in buying and selling prices for ambiguous alternatives compared to risky ones (Eisenberger and Weber 1995). Sonnemann et al. (2009) demonstrate that prediction markets, which were introduced to you in Section 12.6.2, are influenced by the phenomenon of partition dependence (see Section 7.5.4). Furthermore, a study by Gneezy et al. (2003) illustrates that myopic loss aversion, a concept you will learn about in Section 13.3.4.3, is also reflected in market prices.

At this point, we would like to put an end to the discussion on whether or not modeling alternative descriptive preferences is relevant; the opinion (shared by the

authors) that systematic deviations from utility theory are highly relevant when explaining economic relationships seems to prevail.

13.3 Descriptive theories

In Figure 13-1, we have outlined the process of structured and rational decision making from the point of view of utility theory and pointed out that descriptive deviations from the normative benchmark can happen at very different stages of the process. We will also keep an eye on this structure and the phenomena that have been discussed in Sections 13.2.1 to 13.2.4 during the presentation of descriptive theories below. First, we will explain the support theory of Tversky and Koehler (1994) Tversky and Koehler (1994) Tversky (1994) in Section 13.3.1. This theory was developed to model phenomena that occur during the estimation of subjective probabilities. Some of these problems have already been presented in Section 13.2.1. Support theory is no preference theory and hence makes no predictions regarding how individuals will decide. Instead, the goal of support theory is to explain why and how the probability estimates of a decision maker – the inputs to the subsequent evaluation of alternatives – are affected by the framing of a decision problem. All of the following models of support theory are descriptive preference theories. In particular, we will discuss the cumulative prospect theory (CPT) of Tversky and Kahneman (1992), which will be presented in Section 13.3.4. It is undoubtedly the most prominent descriptive decision theory of today and can explain decision-making behavior in a wide range of applications. Before getting to grips with CPT, we will have a look at the original prospect theory (OPT) of Kahneman and Tversky (1979) and rank-dependent utility theories as these older approaches already make use of concepts that reappear in CPT. The last Section, 13.3.5, will present other preference theories that differ even more strongly from expected utility theory than CPT.

The modifications of expected utility theory can have very different starting points. Empirical studies, the mathematically formal weakening of the axioms and insights into decision makers' preferences – as expressed by the paradoxes – give rise to possible modifications of utility theory. In the following highly compressed presentation of preference models, the focus lies on the results of the new theories, i.e. the functional forms of the "new utility functions." The axiomatic underpinnings of the respective theories are only presented when they are essential to understanding. Illustrating the intuition explicitly or implicitly contained in a theory is key (see Machina 1987, Camerer 1995, Diedrich 1997, Starmer 2000, as well as Weber and Camerer 1987). To clarify the various departures from expected utility theory, we will present indifference curves in the probability triangle for most theories (compare Camerer 1989).

13.3.1 Support theory

Support theory is designed to model phenomena related to the estimation of subjective probabilities, as presented in Section 13.1.1. The main idea behind support theory (Tversky and Koehler 1994 as well as Rottenstreich and Tversky 1997) is that decision makers do not assign probabilities to random events but to descriptions of these random events. As a consequence, different descriptions of the same event can lead to different assessments of the probability of occurrence. The decisive factor for the evaluation is the "support" of an event that can be influenced by various heuristics (availability, representativeness, anchoring, and adjustment, etc.). When the description becomes more extensive – think of the empirical investigation of Fischhoff et al. (1978) in Section 13.1.1 – the decision maker will attribute a higher probability to the event. The importance of this theory thus results from the fact that we learn more about how decision makers arrive at their subjective probability estimates. A subjective probability estimate can hence not only differ across individuals but alternative descriptions of the same random event can also have a systematic impact on the estimate. In the following, we want to present the theory in greater detail.

Our following considerations are based on the disjoint events A', B', and their descriptions A, B, referred to as hypotheses in support theory. An event can have more than one description. The event A' - two rolls of a die showing one and two - can be characterized by hypothesis A: "The sum of the points is equal to three," but also by: "The product of points is equal to two." The probability of hypothesis A occurring but hypothesis B not occurring is denoted as $p(A,B)$. Hypothesis A is called the "focal hypothesis." In the traditional thinking of probability calculus this is equivalent to the expression $p(A' \mid A' \cup B')$. In support theory it holds that:

$$\mathrm{p}(A,B) = \frac{s(A)}{s(A)+s(B)}, \tag{13.1}$$

where $s(A)$ denotes the extent of the "support" of hypothesis A on a ratio scale and $s(B)$ the corresponding support of hypothesis B. Now, the question of how the support of a hypothesis is defined becomes crucially important. Here, support theory assumes that for the function s it holds that:

$$s(A) \le s(B \vee C) \le s(B)+s(C), \tag{13.2}$$

given that $A' = (B \vee C)'$ and B and C are disjoint hypotheses. The symbol $\vee$ remains to be clarified. In the definition $A' = (B \vee C)'$, this symbol refers to the fact that the disjoint events B' and C' together characterize the same overall event as does A' while in the description both sub-events are explicitly mentioned. The function s then assigns greater support to the explicit decomposition than to description A which does not mention explicitly the two sub-events. According to (13.2), the support would be even greater if probabilities were elicited separately for both sub-events. Let us have a look at an illustrative example.

Think of your favorite soccer team, e.g. 1899 Hoffenheim (the authors intentionally refrain from considering Alemannia Aachen or Preußen Münster). The event is defined as seeing Hoffenheim lose on the next matchday. Hypothesis A is: "Hoffenheim loses on the next matchday," hypothesis B: "Hoffenheim loses by one goal on the next matchday", and hypothesis C: "Hoffenheim loses by more than one goal on the next matchday" The left part of the inequality in (13.2) means

that the support for losing the match is estimated to be greater when the event is described as "Hoffenheim loses on the next matchday by one goal or loses by more than one goal."

By the use of the complementary hypothesis D: "Hoffenheim will not lose next matchday," the different supports for each description also lead to different probability estimates $p(A, D) \le p(B \vee C, D)$. This more extensive description is termed the "*unpacking* principle"and its consequences have been confirmed by many studies (see Tversky and Koehler 1994). From the right part of the inequality in (13.2) it follows that the support of the two isolated hypotheses B and C in sum would be greater than the support of a combined hypothesis $B \vee C$. It can then be shown that the sum of probabilities $p(B, C \vee D)$ and $p(C, B \vee D)$ also exceeds $p(B \vee C, D)$ (see Rottenstreich and Tversky 1997). If you were thus to ask for the probability of Hoffenheim losing by one goal and afterwards ask for the probability of Hoffenheim losing by more than one goal, the sum of probabilities would be even greater. Support theory is able to explain typical judgment phenomena which you already became acquainted with in Sections 13.2.1 and 13.2.5, e.g. the splitting bias, event splitting, or partition dependence.

13.3.2 Rank-dependent utility theories

When discussing the Allais paradox in Section 13.2.4, we recognized that the linearity of probabilities as claimed by the expected utility formula is a very strong postulate that does not reflect intuitive decision making. *Linearity in probabilities* means that the probabilities of occurrence are used as weighting factors for the utility of the consequences. The 1% difference between 30% and 31% cannot have a different impact than the 1% difference between 0% and 1% or 99% and 100% has. Intuitively, individuals will perceive the change from *almost certain* (99%) to *absolutely certain* (100%) as being much more significant than the change from *possible* (30%) to *a little bit more probable* (31%). Allowing a transformation of probabilities by a probability weighting function in addition to the transformation of the consequences by a utility function suggests itself as a solution. This would yield a value function of the following form:

$$V(a) = \sum_i \pi(p_i) \cdot u(a_i) \tag{13.3}$$

The unequal perception of identical probability differences could then be represented by an appropriately chosen probability weighting function π. The phenomenon given above – that the change from 0% to 1% and the change from 99% to 100% are more important to the decision maker than the change from 30% to 31% – may be explained by a probability weighting function π with $\pi(0\%) = 0\%$, $\pi(1\%) = 3\%$, $\pi(30\%) = 30\%$, $\pi(31\%) = 31\%$, $\pi(99\%) = 97\%$, and $\pi(100\%) = 100\%$.

Unfortunately, approaches as described by (13.3) suffer from serious methodological disadvantages. The sum of the probability weights may be influenced greatly by the decomposition of events into partial events with smaller probabilities; in particular, the sum of the weights does not have to be equal to 1. This may lead to nonsensical preference patterns as it may easily happen that a preference for a stochastically dominated alternative is predicted.

To avoid this problem, a slightly modified form of probability weighting has to be adopted. Not the probabilities of occurrence of singular consequences but rather the cumulated probabilities are to be transformed. *Rank-dependent utility theories* (RDEU theories) constitute such an approach based on the transformation of the entire distribution.

Let the consequences of an alternative a be indexed in ascending order according to the preference of the decision maker: $a_1 \prec \ldots a_i \prec \ldots a_n$. If the preferences fulfill the (yet to be specified) axioms of RDEU, they can be represented by the rank-dependent utility function RDEU(a):

$$\text{RDEU}\,(a) = \sum_{i=1}^{n} u(a_i)\, w(p_1, \ldots, p_i) \tag{13.4}$$

$$w(p_1, \ldots, p_i) = g\Big(\sum_{j=1}^{i} p_j\Big) - g\Big(\sum_{j=1}^{i-1} p_j\Big). \tag{13.5}$$

The arguments of the probability weighting function g are cumulated probabilities. It is assumed that $g(1) = 1$ and $g(0) = g(p_0) = 0$. The $w(p_1, \ldots)$ are called probability weights. For $g(p) = p$ follows $w(p_1, \ldots, p_i) = p_i$, and RDEU is equal to EU. This extension of utility theory is called rank-dependent theory because the strength of the probability transformation for a consequence depends on its rank within the consequences. In particular, the probability weights of two consequences that occur with identical probability are allowed to differ. We agree with you if you believe that this calculus does not look very handy and that it cannot be assumed that the preferences of decision makers are formed by such probability transformations in reality; however, the predictions made by this calculus conform to actual decision behavior surprisingly well (in Diecidue and Wakker (2001), you can read why RDEU theories are much more intuitive than they appear at first glance).

Rank-dependent theories were developed to represent phenomena like the Allais paradox (13.3.4). We therefore want to use this specific example to illustrate the valuation of utility for rank-dependent utility functions. The decision maker is confronted with the situation presented in Figure 13-1, the Allais paradox: a=€3,000, b=(€4,000, 0.8; €0, 0.2), a'=(€3,000, 0.25; €0, 0.75) and b'=(€4,000, 0.2; €0, 0.8). The function g of the decision maker is given in Figure 13-5. It holds that $g(0)=0$, $g(0.2)=0.4$, $g(0.25)=0.45$, $g(0.75)=0.86$, $g(0.8)=0.88$ und $g(1)=1$. To keep things simple, we assume the utility function of the decision maker to be linear, thus u(€0)=0, u(€3,000)=0.75 and u(€4,000)=1. It holds that:

$$\begin{aligned}
\text{RDEU}(a) &= u(3{,}000) = 0.75 \\
\text{RDEU}(b) &= u(0)\,g(0.20) + u(4{,}000)\,(g(1) - g(0.20)) \\
&= 0 + 1 \cdot 0.60 = 0.60
\end{aligned}$$

$$\begin{aligned}
\text{RDEU}(a') &= u(0)\,g(0.75) + u(3{,}000)\,(g(1) - g(0.75)) \\
&= 0 + 0.75 \cdot 0.14 = 0.105 \\
\text{RDEU}(b') &= u(0)\,g(0.80) + u(4{,}000)\,(g(1) - g(0.80)) \\
&= 0 + 1 \cdot 0.12 = 0.12
\end{aligned}$$

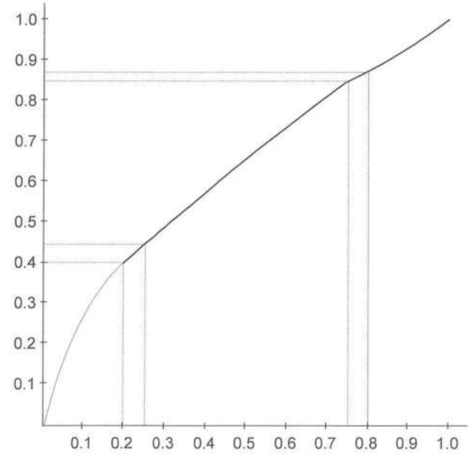


Figure 13-5: Probability weighting function g

This example describes the preference of a decision maker who behaves according to the Allais paradox, i.e. $a \succ b$ and $b \succ a'$.

If the function g is concave, the less preferred consequences are weighted disproportionately high. Accordingly, for g convex, the more preferred consequences are given more weight. Through the transformation of cumulated probabilities, the RDEU theories possess the pleasant property that the sum of all the probability weights is always equal to 1 (= $g(1)$). This is true for other cumulated probabilities as well, for instance if p denotes the cumulated probability that consequence a is not exceeded, the sum of the probability weights for this and all smaller consequences is always equal to $g(p)$ regardless of the probabilities of the smaller consequences occur. A scenario in which the sum of the probability weights is increased by the decomposition of an event cannot emerge. The RDEU theories hence cannot produce violations of stochastic dominance despite the probability transformations.

Rank-dependent utility functions are based on an axiomatic system that is closely related to utility theory. The preferences have to conform to completeness and continuity. Two ways of weakening utility theory's independence axiom lead to rank-dependent utility theories. Firstly, the axiom itself may be weakened to allow a broad class of behavior (on this note, see Quiggin 1982). Secondly, the validity of the axiom may be restricted to a certain class of alternatives (on this note, see Yaari 1987).

The indifference curves of rank-dependent utility functions can also be displayed in a *probability triangle.* Figure 13-6 displays typical RDEU indifference curves for a risk-averse decision maker. These always have a positive slope but are not straight lines as in EUT; they therefore do not conform to the *betweenness* property outlined in Section 13.2.4. If you drew the lotteries of the Allais paradox in the three-outcome diagram, you would recognize that this theory is able to explain the paradox (the example is based on a linear utility function and a $g(p)=p^{0.8}$).

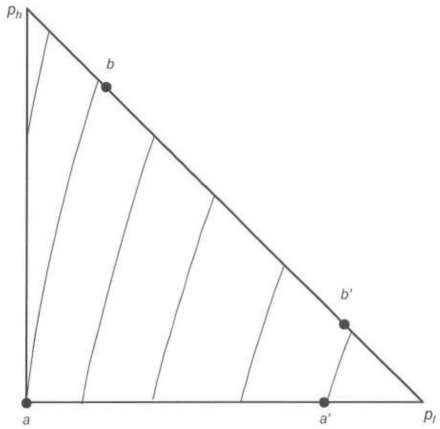


Figure 13-6: Indifference curves for rank-dependent utility functions

Choquet expected utility theory

Rank-dependent utility theories can explain the Allais paradox from Section 13.2.4. However, they cannot suggest an answer to the Ellsberg paradox from Section 13.2.3. This paradox attacks subjective expected utility theory on an even more basic level, namely the postulate that subjective probabilities have to be assigned to each of the states that then have to be used consistently in the preference calculus. In the Ellsberg paradox, we observed a decision behavior that cannot be brought into agreement with any subjective probability distribution imaginable. However, there is an equivalent to RDEU, the *Choquet expected utility theory* (CEU theory), which is capable of representing decision patterns such as in the case of the Ellsberg paradox. We will look at the approach of Schmeidler (1989) (see also Gilboa 1989 which is a generalization of the subjective expected utility theory by Savage (1954); put more exactly, it is a generalization of the axiomatiza-

tion of the subjective expected utility theory by Anscombe and Aumann 1963). Just as with Savage, consequences which occur depending on environmental states are assigned to each alternative. If the SEU's *sure thing principle* is only required for co-monotone pairs of alternatives, one obtains the CEU theory. Two alternatives a and b are co-monotone if there are no states $s_i, s_j \in S$ with $a(s_i) > a(s_j)$ and $b(s_j) > b(s_i)$; this means that there must not exist a pair of states for which the consequence of one alternative improves by the change of one state to the other while the consequence of the other alternative worsens due to the very same change in states.

For the expected utility of the Choquet utility theory, it holds that:

$$\mathrm{CEU}(a) = \sum_{i=1}^{n} u_i \, [q(\bigcup_{j=i}^{n} s_j) - q(\bigcup_{j=i+1}^{n} s_j)] \quad \text{with } u_i = u(a(s_i)), \quad u_1 \leq u_2 \leq \ldots u_n. \tag{13.6}$$

Of the functions appearing, you know the function u (the von Neumann and Morgenstern utility function) from Chapter 9. The function q is a real-valued function that represents a (not necessarily additive) probability measure. To distinguish additive from non-additive probabilities, we will refer to non-additive probabilities as capacities. The capacities possess the following properties:

- $q(S)=1$ with $S=$ set of all states
- $q(\emptyset)=0$
- $q(T) \leq q(T')$, if $T \subseteq T'$.

The capacities of a disjoint decomposition of S (e.g. an event and its complement) may sum up to a number greater or smaller than 1. If the sum is greater (smaller) than 1, the decision maker is ambiguity affine (ambiguity averse). The degree of the attitude towards ambiguity is measured by the degree of super- or subadditivity of the capacities in Choquet utility theory. If the capacities are additive, the CEU is identical to the SEU.

The formula for the Choquet utility strongly resembles the calculation of the rank-dependent utility. Analogously, in order to calculate the CEU value, the states have to be sorted into descending (attention, this is the opposite from RDEU!) order according to the utility of the consequences occurring for each alternative. The CEU value can then be calculated analogously to the RDEU value for rank-dependent utility functions.

To gain an intuition for CEU theory, we would like you to compare the following two urns:

Urn 1:

The urn contains 10 white and 10 yellow balls. You win €100 if a white ball is drawn. You receive €0 if a yellow ball is drawn.

Urn 2:

The urn contains 20 white and yellow balls. You can determine your winning color. You receive €100 if your winning color is drawn and €0 otherwise.

Urn 1 corresponds to a simple 50/50 lottery whereas urn 2 represents an ambiguous lottery. The subjective expected utility theory cannot differentiate between the two urns and predicts that the decision maker will be indifferent between the urns. (If you have the "feeling" that there is a majority of the balls of one color in urn 2, then you even should select urn 2 because your probability of winning is greater than with urn 1.)

CEU theory predicts that an ambiguity-averse decision maker will prefer urn 1 over urn 2. While the capacities are similar to the probabilities for urn 1, i.e. $q(\text{white}) = 0.5$ and $q(\text{yellow}) = 0.5$, the capacities for urn 2 may well be different from 0.5. Empirical research shows that capacities for ambiguous alternatives analogous to urn 2 can be found in the area of 0.4 (see Keppe and Weber 1995 as well as Mangelsdorff and Weber 1994): the CEU utility of urn 2 is therefore usually less than the CEU utility of urn 1.

13.3.3 Cumulative Prospect Theory

Cumulative Prospect Theory (Tversky and Kahneman 1992) is an extension of the original Prospect Theory by Kahneman and Tversky (1979) and is currently the most prominent descriptive preference theory. Nowadays, when you hear or read something in the literature (also in this book) on *Prospect Theory*, you should assume that it will refer to the cumulative version. The original Prospect Theory (OPT) from 1979 is only of historical importance today. However, to prevent possible misunderstandings, the cumulative version of Prospect Theory is commonly referred to as CPT.

Thinking back to the problems discussed in Section 13.2, CPT is able to model both the reference point effects of Section 13.2.3 and the Allais paradox of Section 13.2.4. This is accomplished by changing the assumptions regarding the value function and by introducing probability weighting of the kind we already know from the RDEU theories.

13.3.3.1 Reference dependent evaluation

A major difference between Prospect Theory and RDEU theories lies in the *reference dependent evaluation*. The decision maker does not use final wealth as an argument in his value function but instead evaluates changes relative to a reference point perceived as neutral. Let us look at an example. The decision maker has a wealth of €50,000 and has to evaluate the risky alternative (€100, 0.5; –€100, 0.5). Under expected utility theory, this lottery must be presented as a lottery over final wealth levels. The lottery (€50,100, 0.5, €49,900, 0.5) is hence to be evaluated. From the perspective of Prospect Theory, the decision maker would evaluate the lottery with respect to a reference point, in the example probably his with respect to his current wealth of €50,000. During his evaluation he would thus effectively think about the lottery (€100, 0.5, –€100, 0.5).

The status quo is generally perceived as the reference point, but other reference points are also possible. If you have received a pay raise of 3% every year, you will eventually perceive a 2% raise as a loss even if it is an improvement relative

to the status quo (your reference point is then the expected increase by 3%). The coding of consequences relative to the reference point is included in the so-called editing phase of Prospect Theory that precedes the actual evaluation. The editing phase was much more important for the original version of Prospect Theory (OPT) and besides *coding* included the "cleaning-up measures" (*aggregation, simplification, combination*). Today, these editing operations are hardly discussed. We also want to focus on coding as the most important preparatory step for a reference-dependent evaluation.

In Prospect Theory, the (coded) consequences of the alternatives are evaluated by use of a measurable value function. In general, this value function can be elicited with the procedures presented in Chapter 5. When it comes to terminology, we have to be careful here. As Prospect Theory was mainly developed for the evaluation of risky alternatives, the curvature of the value function is commonly interpreted in terms of risk aversion and risk proclivity. Strictly speaking, this is incorrect as it refers to diminishing or increasing value sensitivity (see Abdellaoui et al. 2007). Overall, the analysis of risk attitudes within Prospect Theory is much more complicated than under expected utility theory because the evaluation of risky lotteries in PT depends on multiple factors (diminishing value sensitivity, loss aversion, probability weighting). In Section 13.3.3.4, there will be more details on this issue.

Based on empirical findings, Kahneman and Tversky assume the value function of a decision maker to be convex over gains and concave over losses. Furthermore, the value of a gain is smaller than the absolute value of an equally sized loss (loss aversion). This asymmetry concerning gains and losses allows for the modeling of reference point effects. At the same time, it highlights the importance of the setting of the reference point. Figure 13-4 illustrates the typical shape of a value function. Before we have a closer look at the properties of this value function, we would like to review the evaluation concept of Prospect Theory in general.

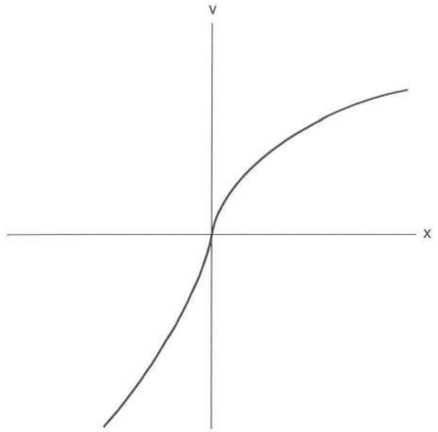


Figure 13-7: The value function of Prospect Theory

13.3.3.2 The evaluation concept of CPT

In CPT, the consequences $a_i (i = 1, ..., n)$ of an alternative are sorted in ascending order, just as in RDEU. The reference point of a decision maker is subsequently used to split up the set of consequences into gains and losses. We assume that our alternative m has negative and $n - m$ positive consequences. (The case where one of the consequences is equal to zero can easily be ignored because it does not matter whether we treat such a consequence as a gain or a loss – more on this later). Gains and losses are then evaluated with the help of the value function as in original Prospect Theory. The decision weights w_i are also determined *separately* for gains and losses. This is an important difference to the approach of RDEU. The reference point is the starting point for all classifications under CPT – there are gains in the one direction and losses in the other. Conceptually, the changes in both directions are always treated very similarly. This is not only reflected in the value function of Prospect Theory (familiar to you from Figure 13-7) that exhibits diminishing value sensitivity in both directions when starting from the reference point (i.e. for gains and losses). Such "symmetry" can also be observed for the probability weighting function, especially with reference to the question of the direction in which the probabilities are to be cumulated before they are transformed. CPT gives the following answer: in no absolute direction but always from the large changes to the smaller ones. This not only requires a separate accumulation procedure for gains and losses, it also requires an accumulation of gains from the largest values to the smaller ones, whereas losses are cumulated from the smallest

values to the larger ones (which in absolute terms also corresponds to the direction of large to small values). This (reference) point symmetric approach leads to a somewhat unhandy notation. We will first write down the formula explicitly and then illustrate it with the help of an example.

In CPT, the evaluation of a risky alternative denoted by CPT(a) is calculated as the sum of the expected rank-dependent utility of the positive as well as the negative consequences by the following formula:

$$\mathrm{CPT}(a) = \sum_{i=1}^{m} v(a_i)\, w^-(p_i) + \sum_{i=m+1}^{n} v(a_i)\, w^+(p_i)\,. \tag{13.7}$$

The decision weights of the positive $w^+(p_i)$ and the negative consequences $w^-(p_i)$ are calculated separately according to the following formulas:

$$w^-(p_i) = g^-\left(\sum_{j=1}^{i} p_j\right) - g^-\left(\sum_{j=1}^{i-1} p_j\right) \tag{13.8a}$$

$$w^+(p_i) = g^+\left(\sum_{j=i}^{n} p_j\right) - g^+\left(\sum_{j=i+1}^{n} p_j\right) \tag{13.8b}$$

The different formulas for calculating the probability weights for gains and losses do not result from differences in the overall approach but are simply due to fact that consequences with the highest index represent the largest changes for gains, whereas the consequences with the lowest index lead to the largest changes for losses. When you think of all consequences in terms of the line of real numbers (with the reference point somewhere in the middle), the rule for accumulating always is "from the outside inwards". We hence accumulate from right to left for gains and from left to right for losses. We want to illustrate the evaluation concept of CPT with the help of a sample calculation. For the value function, we use the two-part power function as proposed by Tversky and Kahneman (1992) in their original work:

$$v(x_i) = \begin{cases} x_i^{0.88} & \text{for } x_i \geq 0 \\ -2.25 \cdot (-x_i)^{0.88} & \text{for } x_i < 0 \end{cases} \tag{13.9}$$

that also forms the basis for Figure 13-7.

To keeps things simple, we additionally assume the probability weighting for positive outcomes $g^+(p_i)$ to be equal to the probability weighting for negative outcomes $g^-(p_i)$. For the probability weighting function g there also exist (in part quite complex) functional forms that are discussed in the literature (see Section 13.3.3.3). However, we will just provide you with some values in a table so that the key aspect of the example (the process of probability transformation) is explicitly highlighted. Table 13-1 gives you these values for $g^+(p_i)=g^-(p_i)=g(p_i)$ and Figure 13-8 provides a graphical representation.

Table 13-1: Selected values for the probability weighting function g

p_i	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$g(p_i)$	0.00	0.08	0.14	0.19	0.23	0.27	0.31	0.35	0.39	0.42	0.45

p_i	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
$g(p_i)$	0.49	0.53	0.57	0.62	0.67	0.73	0.79	0.85	0.92	1.00

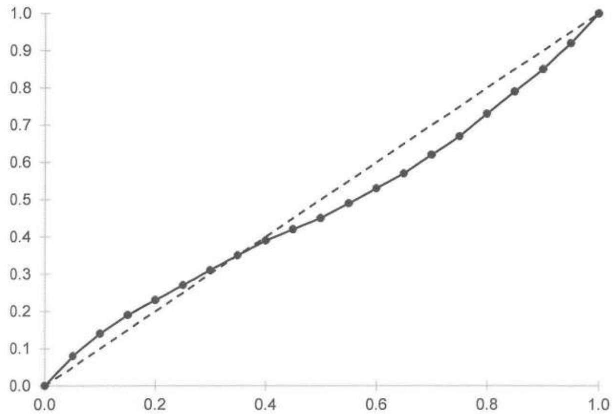


Figure 13-8: The probability weighting function g as used in the example calculation

To calculate the transformed probabilities w_i, three steps need to be performed:

- Accumulation of the initial probabilities
- Transformation of the cumulated probabilities
- Decumulation of the transformed cumulated probabilities

In Table 13-2, these steps are performed for a lottery with 8 consequences. The lottery itself can be identified from rows (3) and (4). It has four negative and three positive outcomes. The valuation of the outcomes that results from simply inserting into the value function (13.9) is given in row (2). According to the formulas (13.8a) and (13.8b), both accumulating and decumulating have to be performed "from the outside inwards". In each case, we start with the positive and negative outcome that is largest in absolute terms.

Table 13-2: Process of probability transformation in CPT

(1)	i	1	2	3	4	5	6	7	8
(2)	$v(a_i)$	−14.02	−7.62	−4.14	−2.25	0	+4.84	+6.91	+8.90
(3)	a_i	−8	−4	−2	−1	0	+6	+9	+12
(4)	p_i	5%	10%	5%	40%	10%	15%	5%	10%
(5)	cumul. p_i	5%	15%	20%	60%	–	30%	15%	10%
(6)	transf. $\sum p_i$	8%	19%	23%	53%	–	31%	19%	14%
(7)	$w(p_i)$	8%	11%	4%	30%	–	12%	5%	14%
(8)	$v(a_i) \cdot w(p_i)$	−1.12	−0.84	−0.17	−0.68	–	0.58	0.35	1.25

First, row (5) presents the accumulation of probabilities. For losses ($i = 1, ..., 4$), the accumulation occurs from left to right. For gains ($i = 8, ..., 6$), it occurs from right to left. Second, in row (6), the accumulated probabilities are transformed using the function g (see Table 13-1). Third, and last, in row (7) it is again decumulated from the outside inwards. You can think of the principle of decumulation in the sense that accumulating row (7) must yield row (6). As a result, the probability weights have been determined in row (7). Drawing on these weights, the "expected valuation" (the sum of entries in row (8)) results in a CPT-value of

$$\sum_{i=1}^{8} v(a_i) \cdot w(p_i) = -0.63 < 0 .$$

The exemplary CPT decision maker would reject participating in the lottery. The risky alternative holds a CPT value below the CPT value of 0 that would result from not participating in the lottery and keeping up the status quo.

On the basis of this example and Table 13-2, we would like to point out some important aspects of CPT evaluation. On the one hand, you can see that during probability transformation we did not pay any attention to column 5 that contains consequence $a_5 = 0$. This is so because the consequence, due to its evaluation of $v(0) = 0$, does not provide any contribution to the overall CPT evaluation of the alternative. Consequently, it does not matter whether we count the consequence 0 towards the positive or the negative outcomes (a slightly different probability weight of 8% and 9%, respectively, would result, but this is irrelevant for the evaluation). If you remember this fact concerning the irrelevance of the consequence 0, you also have a memory hook in case you forget about whether the accumulation needs to be occur from the outside in or from the inside out: in case we were to accumulate inside out, it would very well matter whether the consequence 0 is added to the positive or the negative consequences. In such a case, all outside consequences would be affected by the (probability of the) zero consequence when accumulating. Only because we accumulate outside in, the consequence 0 can be completely disregarded as it would only be added in the last step of accumulating anyway.

As a second interesting aspect, it can be observed that the sum of all probability weights does not add up to 1 (even if you had included the probability weight of the consequence 0). At a first glance, this may be irritating to you, as we argued in Section 13.3.2 based on just this weighting phenomenon (the sum of weights being different from1) that violations of stochastic dominance can result from a direct transformation of probabilities. However, we can reassure you that the accumulated probability weighting of CPT eliminates violations of stochastic dominance even if the sum of the weights generally does not add up to 1for mixed lotteries, i.e. lotteries with both positive and negative outcomes.

Finally, it can clearly be seen from Table 13-2 that a probability can be perceived very differently within CPT depending on the rank of the associated consequence. Columns 1, 3 and 7 show that a (objective) probability of 5% at times can enter the CPT evaluation with a weight of 8%, another time with 4%, and yet another time with 5%. It is certainly not true that small probabilities tend to be overweighted in CPT evaluation as some "simplified" descriptions of Prospect Theory suggest (or as suggested by outdated descriptions as this holds true for Original Prospect Theory). The type of transformation a small probability is subjected to is rather determined by where it is "inserted" in the cumulative distribution. Still, however, the probability weighting is by no means done arbitrarily. There indeed exist systematic distortions within probability weighting which we will now have a closer look at.

13.3.3.3 CPT's probability weighting and value functions

Probability weighting

CPT's typical functional forms for probability weighting g have a shape as presented in Figure 13-9 (see Tversky 1992). These functions are continuous, extend from 0 to 1 (more precisely: $g(0) = 0$ and $g(1) = 1$), and have an inverse S-shape. They are thus concave for small and convex for large probabilities. It is absolutely necessary to keep in mind that the functions in Figure 13-9 transform the cumulative distribution and not the individual probabilities. In the case that for the weighting function it holds that $g(2\%) = 10\%$, a probability of 2% is only strongly overweighted at 10% when it belongs to one of the two most extreme outcomes (largest gain or loss). Only then is the accumulated probability (up to this point) equal to the individual probability and thus the probability transformed correspondingly strongly. If the probability of 2% instead belongs to a consequence that increases the accumulated probability from 35% to 37%, its weight results from the difference between $g(35\%)$ and $g(37\%)$. This difference, however, can be very small due to the flatness of the curve in this region, perhaps amounting to only 0.5%.

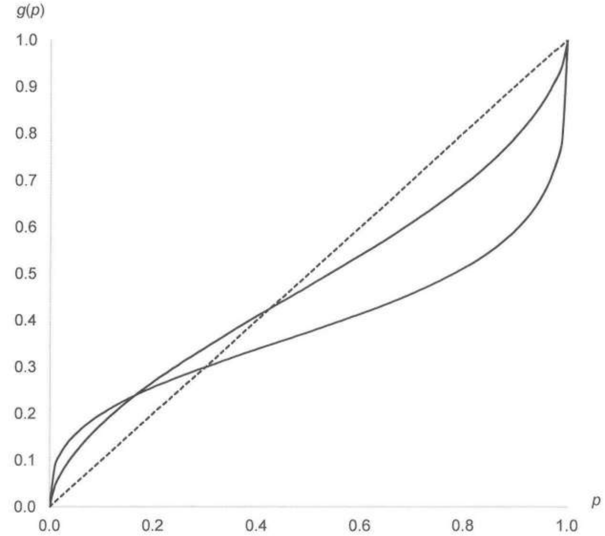


Figure 13-9: The typical shape of the probability weighting function

These considerations depict the basic pattern of probability weighting: due to the inverse S-shape of the function g, not the small individual probabilities but the probabilities associated with the most extreme outcomes are overweighted. The probabilities associated with moderate outcomes (medium-sized gains and losses) are underweighted instead. Very extreme overweighting of probabilities can be observed if the most extreme outcomes occur with a very small probability; this is the typical lottery phenomenon. The probability of winning the German lottery jackpot amounts to 0.000000715%. If this probability is strongly overweighted when evaluating the purchase of a lottery ticket, indeed a positive CPT value of playing the lottery can result.

Regarding the functional form of probability weighting, a variety of approaches is discussed in the literature. The form as presented in Equation (13.10) was proposed in Tversky and Kahneman's (1992) original paper and also calibrated in an experimental study.

$$g(p) = \frac{p^{\delta}}{(p^{\delta} + (1-p)^{\delta})^{1/\delta}} \tag{13.10}$$

The curvature of the function and thus the strength of probability weighting is determined by the (single) parameter δ. For $\delta < 1$, we obtain the characteristic inverse

S-shape. The smaller δ becomes, the more pronounced the curvature of the function is (for $\delta = 1$, there is no probability weighting). Tversky and Kahneman (1992) determined an average of $\delta = 0.61$ for the weighting of gain probabilities and an average of $\delta = 0.68$ for the weighting of loss probabilities. Another commonly applied single-parameter functional form is proposed by Prelec (1998). It stands out from the rest because it can be motivated by axiomatic considerations.

Aside from the single-parameter functional specifications, two-parameter forms are also discussed in the literature. The class of functions advocated by Gonzalez and Wu (1999)

$$g(p) = \frac{\delta \cdot p^{\gamma}}{(\delta \cdot p^{\gamma} + (1-p)^{\gamma})} \tag{13.11}$$

is based on the idea that the probability weighting function demonstrates two distinct psychological effects. One aspect, termed *discriminability*, reflects the actual differences in the perception of probabilities at different points of the cumulative distribution and is captured by the *curvature* of the function. The other effect, namely *attractiveness*, refers to a general over- or underweighting of probabilities and influences the *elevation* of the function. As shown in Figure 13-10, curvature and elevation can be controlled independently of each other via the parameters γ and δ for the functional type (13.11). The left panel of the figure presents a distribution with less curvature ($\gamma = 0.7$ and $\delta = 0.9$) in addition to the base case ($\gamma = 0.5$ and $\delta = 0.9$) in bold line. The right panel contrasts the base case with an overall less elevated distribution ($\gamma = 0.5$ and $\delta = 0.5$). The two-parameter specification thus allows for an even more differentiated calibration of individual probability weighting.

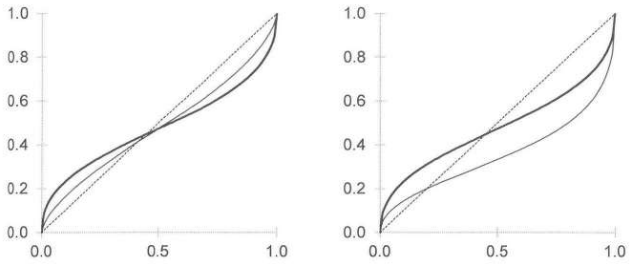


Figure 13-10: Probability weighting functions by Gonzalez und Wu (1999)

The value function

The value function v of prospect theory is characterized by three main properties that are commonly perceived as integral parts of the theory. They have all been mentioned before but we would like to use this opportunity to recapitulate and to discuss them in some more detail:

(a) The value function of prospect theory evaluates changes relative to a reference point. Positive values are assigned to positive changes with respect to the reference point, negative changes receive negative values. The reference point itself is evaluated with 0 (i.e. $v(0) = 0$).

(b) The value function reflects diminishing value sensitivity for both gains and losses; meaning that it is convex in the domain of gains and concave in the domain of losses.

(c) The value function is steeper in the domain of losses than in the domain of gains, i.e. in absolute terms, a loss receives a higher valuation than an equally sized gain. This property is referred to as *loss aversion*. The value function has a *kink* at the reference point.

For the functional form of the value function, a power function is commonly used. This form was also proposed by Tversky and Kahneman (1992) and became the standard thanks to its manageability and easy interpretability. In its general form:

$$v(x_i) = \begin{cases} x_i^{\alpha} & \text{for } x_i \geq 0 \\ -k \cdot (-x_i)^{\beta} & \text{for } x_i < 0 \end{cases} \tag{13.12}$$

which we have already used in (13.8); the curvature for gains, the curvature for losses, and the degree of loss aversion can be controlled separately via the parameters α, β, and k. When all three parameters have a value of 1, the value function is linear and the decision maker evaluates the alternatives according to their expected value (unless he conducts probability weighting). The case of $\alpha = \beta = 1$ and $k > 1$ is also often considered in the literature; the value function then is piecewise linear with a kink at the reference point and only reflects loss aversion. This simplification is carried out because the loss-aversion effect is rated as much more stable and relevant than diminishing value sensitivity.

Nevertheless, it is generally assumed that not only $k > 1$ but also $0 < \alpha, \beta < 1$ hold. We then obtain the typical form of the prospect theory value function with a curvature that changes from convex to concave at the reference point and that is steeper in the domain of losses. Tversky and Kahneman (1992) present an experimental study where 25 students evaluate 64 lotteries. Based on the responses of the students, they calculate the best-fitting prospect theory parameters. The combination $\alpha = 0.88$, $\beta = 0.88$, and $k = 2.25$ (together with the parameters $\delta^{+} = 0.61$ and $\delta^{-} = 0.69$ for probability weighting according to (13.10), which they obtain as the median of their estimation) is nowadays used as the standard for prospect theory parameters. We will also imply these values when we speaking of a *standard Prospect Theory decision maker* .

Subsequently, many additional studies have been carried out that estimate prospect theory parameters with improved methods and more extensive populations. The results, however, are not entirely unambiguous. While some studies have found that the curvature is less pronounced in the loss domain (the average β is thus closer to 1), some newer studies documented just this effect for the curvature in the gain domain (for an overview, see Wakker et al. 2007). The loss-aversion effect ($k > 1$) is nevertheless stable across all studies, at least on an aggregate level (see e.g. Köbberling and Wakker 2005 or Abdellaoui et al. 2007). Individual data, however, contain much noise. Many studies have analyzed whether the general parametric form of the value function can be supported (e.g. Abdellaoui 2000 or Stott 2006). In these studies, it was shown that the power function indeed represents a very well-suited functional class.

13.3.3.4 Risk attitudes within CPT

In expected utility theory, the risk attitude of a decision maker is exclusively determined by the curvature of his/her utility function. The decision maker acts as risk averse if the utility function is concave, i.e. for every lottery he/she assigns a certainty equivalent that is smaller than the expected value of the lottery. In prospect theory, things are much more complicated. The interplay of value function and probability weighting function can create very diverse patterns: the risk attitude towards a specific lottery is determined jointly by all the components. The basic definitions of risk aversion, risk neutrality, and risk proneness remain unaffected, however: when evaluating a lottery, a decision maker acts as risk averse if his/her certainty equivalent is below the expected value, which results in a positive risk premium. Accordingly, he behaves as risk neutral or risk loving if the certainty equivalent is equal to or above the expected value, respectively. Consequently, within prospect theory it is also not complicated to determine the risk attitude of a decision maker towards a specific lottery when his/her CPT functions are known. It is however much more complicated to attribute the calculated (positive or negative) risk premium to the specific components of CPT evaluation and to characterize in general the lotteries where in one or the other type of risk attitude will be observed.

In Section 13.2.3 we have already seen in an example that many decision makers prefer a sure gain of €50 to a lottery over gains of (€100, 0.5; €0, 0.5). At the same time, they prefer the lottery over losses (€0, 0.5; –€100, 0.5) to a sure loss of €50. The observation of different risk attitudes, risk aversion for lotteries over gains, and risk seeking for lotteries over losses can easily be explained by the differences in curvature of the PT value function for gains and losses. The convexity in the gain domain leads to risk aversion; the convexity in the domain of losses leads to risk seeking. (In the spirit of Section 9.6.1, it should be discussed at this point whether this indeed reflects differences in risk attitudes of just the effects of differences in value sensitivity but we do not want to pick up this discussion again.) This basic pattern of risk attitudes can be overthrown by probability weighting. Concerning lotteries over gains with a small probability of very large gains, the strong overweighting of the probability of success can possibly overcompensate for the concavity of the value function. We would like to present a clarifying example.

Let us look at a simple lottery over gains of the type $a = (€0, 0.5, €100, 0.5)$. The standard CPT decision maker from Section 13.3.3.3 will assign to this lottery a CPT value of $g(0.5) \cdot v(100) = 0.42 \cdot 57.54 = 24.20$. The certainty equivalent of a is

thus equal to €37.38, as $v(€37.38) = 24.20$. Accordingly, the risk premium $\mathrm{RP}(a) = \mathrm{EV}(a) - \mathrm{CE}(a) = €50 - €37.38 = €12.62$ is positive and the decision maker is risk averse.

If, instead, the lottery over gains is characterized by a small probability of a very large gain, for instance as in $b = (€1{,}000, 0.05; €0, 0.95)$, the risk attitude changes. The standard CPT decision maker from Section 13.3.3.3 will assign to this a CPT value of $g(0.05) \cdot v(1000) = 0.132 \cdot 436.51 = 57.46$. The certainty equivalent of b is equal to €99.83 and the risk premium $\mathrm{RP}(b) = \mathrm{EV}(b) - \mathrm{CE}(b) = €50 - €99.83 = -€49.83$ is negative. The decision maker behaves as risk loving.

The same phenomenon can be observed for losses. The general risk-seeking behavior that is due to the convexity of the value function is reverted to risk-averse behavior by probability weighting in the case of improbable but substantial losses. In Table 13-3 we have portrayed this pattern of risk attitudes - also known as the "*fourfold pattern of risk attitudes*" (see Tversky and Wakker 1995) - augmented with examples. The associated calculations only differ from the ones just presented in the fact that a somewhat smaller curvature of the probability weighting function ($\delta^- = 0.69$) is assumed and that the value function v includes the loss aversion coefficient of $k = 2.25$.

The fourfold pattern of Table 13-3 describes the risk attitude for pure lotteries over either gains or losses. When looking at mixed lotteries where both losses and gains can occur in the same lottery, it becomes increasingly difficult to characterize the lotteries that reveal risk aversion or risk seeking. In addition to probability weighting and the differences in curvature of the value function, loss aversion (the value function running steeper in the domain of losses than in the domain of gains) also needs to be accounted for. In the process, loss aversion will generally emerge as the dominant factor. If losses are perceived negatively by more than twice as much as equally sized gains, lotteries with a positive expected value also often lead to negative CPT values. An example that is widely used in the discussion on loss aversion is the lottery (–€100, 0.5; +€200, 0.5). It holds an expected value of +€50, well above zero; however, a sufficiently loss-averse decision maker will refuse to participate in this lottery: the loss of €100 outweighs the equally probable gain of €200. Our example calculations in Table 13-3 also show that the negative CPT value of the lottery a (with a positive expected value of +1.5) is mainly driven by the pronounced loss aversion whereas probability weighting and diminishing value sensitivity only have a minor impact. Nevertheless, for mixed lotteries examples can also be constructed where strong probability weighting or an especially strong influence of risk seeking in the domain of losses leads to risk-prone overall behavior.

Table 13-3: The fourfold pattern of CPT risk attitudes

Gains		Losses	
"Normal" lotteries over gains Example		"Normal" lotteries over losses Example	
0.5 → €100 0.5 → €0	CPT=+24,20 CE=+€37.38 RP=+€12.62	0.5 → €0 0.5 → –€100	CPT=–58,78 CE=–€24.21 RP=–€25.79
Risk aversion		Risk seeking	
Lotteries over gains with a small chance of a high gain Example		Lotteries over losses with a small chance of a high loss Example	
0.05 → €1,000 0.95 → €0	CPT=+57.46 CE=+€99.83 RP=–€49.83	0.95 → €0 0.05 → –€1,000	CPT=–109.45 CE=–€82.62 RP=+€32.62
Risk seeking		Risk aversion	

CPT indifference curves can also be visualized in a probability triangle. As the shape of these curves is influenced only by probability weighting and not by the value function (only three outcomes with fixed CPT values are considered), the diagram for CPT does not differ substantially from the one for RDEU theory presented in Figure 13-6. The differences in the curvature of the indifference curves result from the diverging assumptions on the functional form of the probability weighting function. The three-outcome-diagram in Figure 13-11 is based on a standard CPT decision maker and a lottery with the outcomes $x_h = €200$, $x_m = €0$, and $x_l = -€100$.

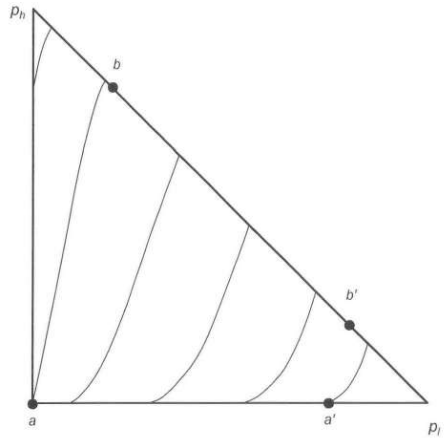


Figure 13-11: Three-outcome-diagram for cumulative prospect theory (CPT)

Loss aversion versus Risk aversion

In general, the strong influence of loss aversion on the evaluation of mixed lotteries (without any doubt, alternatives with both gains and losses are the most relevant to everyday life) leads to risk averse behavior and positive risk premia. It was thus widely discussed whether it is necessary to introduce the additional new concept of "loss aversion". Supporters of the *rational camp* argued that the typical pattern of (largely) risk averse behavior could just as well be explained by expected utility theory and the common assumption of a concave utility function. In addition, the objection of the supporters of the *behavioral camp* that risk tolerance decisions are carried completely independently of other wealth could supposedly be refuted: for a suitable utility function, the fact that a decision maker evaluates a lottery equally for any level of given wealth could also be modeled within the limits of expected utility theory.

This argument was convincingly disproved by Rabin (2000). He showed that the "risk aversion" commonly observed for lotteries with small consequences (that, in fact, is loss aversion which occurs for small consequences due to the kink in the value function) cannot be explained by expected utility theory. In detail, he pointed out that an expected utility maximizer who rejects a 50/50 lottery on a loss of €100 or a gain of €110 independent of his prior wealth level must also reject a 50/50 lottery on a gain of €10M or a loss of €1,000 (a preference that seems highly questionable). The intuition behind his results is as follows: a utility function which in every point is sufficiently concave to explain the strong risk aversion for lotteries with small consequences must have such high curvature that an absolutely unrealistic risk behavior results for large gambles. Rabin's calibration theorem thus proves that the decision pattern observed in reality can indeed be explained by loss aversion but not by risk aversion which (according to expected utility theory) results from the curvature of the utility function.

Elicitation of probability weighting and value functions
The strong interaction between probability weighting, diminishing value sensitivity, and loss aversion not only leads to complex patterns in risk attitude but also complicates inference on the exact shape of both functions from lottery choices or stated certainty equivalents. The simple approaches that we have presented in Chapter 9 for the elicitation of utility functions cannot be applied without modifications. One feasible strategy (see Tversky and Kahneman 1992) is to ask for certainty equivalents of a large set of different lotteries and then to simultaneously estimate the most suitable parameter combination (α, β, k, δ^+, δ^-). However, it can be observed that - due to the pronounced interaction effects - quite different parameter combinations can have a similar explanatory power (see Zeisberger et al. 2010).

An alternative approach is to determine the components of prospect theory (e.g. the curvature of the value function of the domain of gains) independently using procedures that remain unaffected by the other components (e.g. probability weighting). You have already been acquainted with an example of such an approach in Section 9.4.6, namely the trade-off method for utility functions. It identified points on the utility function via lottery comparisons where an identical increase in utility could be inferred for every step in the sequence of consequences even if the probabilities enter the preference with a systematic bias. Similar trade-off methods have been developed for the elicitation of the PT value function (e.g. Abdellaoui et al. 2008). However, they are rather complex and will not be discussed at this point.

13.3.3.5 Prospect theory in real life

Prospect theory is able to describe and explain a variety of phenomena and decision patterns in everyday life. Camerer (2000) collected various applications and reports which component of prospect theory is decisive for explaining the observed behavior. As can be seen from Table 13-4, these examples cover a wide range of areas of application. They extend from finance and job market phenomena to customer behavior and gambling behavior on the racetrack.

We have already talked about the example presented in the last row of Table 13-4 in the section on probability weighting (13.3.3.3). The overweighting of very small probabilities of winning can lead to situations where playing the lottery seems attractive, especially if a very high but also very unlikely jackpot is on offer. The decision maker is then placed in the lower left cell of the four-cell matrix of Table 13-3 and acts as risk loving. Also, the situation that corresponds to the lower-right cell of the four-cell matrix is represented in the examples. The next-to-last row refers to the phenomenon that excessive premia are paid for insurance on very improbable cases of loss. This example of risk aversion in the domain of losses is driven by a strong overweighting of the small probability (of a loss).

Table 13-4: Examples of Prospect Theory in the real life

Domain	Phenomenon	Description	Type of Data	Ingredients	References
Stock market	Equity premium	Stock returns are too high relative to bond returns	NYSE stock, bond returns	Loss aversion	Benartzi and Thaler (1995)
Stock market	Disposition effect	Hold losing stocks too long, sell winners too early	Individual investor trades	Reflection effect	Odean (1999); Genesove and Mayer (2001)
Labor economics	Downward-sloping labor supply	NYC cabdrivers quit around daily income target	Cabdriver hours, earnings	Loss aversion	Camerer et al. (1997)
Consumer goods	Asymmetric price elasticities	Purchases more sensitive to price increases than cuts	Product purchases (scanner data)	Loss aversion	Hardie et al. (1993)
Macroeconomics	Insensitivity to bad income news	Consumers do not cut consumption after bad income news	Teachers' earnings, savings	Loss aversion, reflection effect	Shea (1995); Bowman et al. (1999)
Consumer choice	Status quo bias, Default Bias	Consumers do not switch health plans, choose default insurance	Health plan, insurance choices	Loss aversion	Samuelson and Zeckhauser (1988); Johnson et al. (1993)
Horse race betting	Favorite-longshot bias	Favorites are underbet, longshots overbet	Track odds	Overweight low p(loss)	Jullien and Salanié (2000)
Horse race betting	End-of-the-day effect	Shift to longshots at the end of the day	Track odds	Reflection effect	McGlothlin (1956)
Insurance	Buying phone wire insurance	Consumers buy overprices insurance	Phone wire insurance purchases	Overweight low p(loss)	Cicchetti and Dubin (1994)
Lottery betting	Demand for Lotto	More tickets sold as top prize rises	State lottery sales	Overweight low p(win)	Cook and Clotfelter (1993)

In the following we want to have a somewhat closer look at two other areas of application for prospect theory; both come from the finance domain.

The disposition effect

You are thinking about selling a stock from your portfolio. If the stock performed well after the purchase, you will find the decision easy. Conversely, if the stock is in the red (meaning you would have to sell at a loss), the decision would probably be a much harder one. This phenomenon of often selling winner stocks more readily and faster than losers is referred to as the disposition effect (Shefrin and Statman 1985). It constitutes a "phenomenon" because on rational accounts the (historical) purchase price should not have an influence on the decision to sell a stock (abstracting from any tax issues). Whether is it beneficial to hold or sell should only depend on the expectations about the future development of the stock. However, a strong tendency to sell winners (stocks that are currently quoted above the purchase price) too early and losers (stocks where the current price is below the purchase price) too late can be observed. This was demonstrated by Odean (1998) in an extensive analysis of discount broker data. To rationalize their behavior, investors often state that they hold on to the losers because "they will bounce back" (the belief in *mean reversion*). This is a costly misestimation: Odean finds that, in the following year, the – just sold – winners perform better than the unsold loser stocks by 3.4% – unfortunately, this performance no longer benefits the investors.

But how can the behavioral pattern of the disposition effect be explained by prospect theory? The crucial component is the reference point-dependent evaluation and the differences in curvature in the domains of gains and losses. For simplicity, let us have a look at a stock that is currently quoted at €1,100 and has a 50/50 chance of going up to €1,200 or down to €1,000. Assuming you bought the stock at €1,000 some time ago, you hold a "winner stock" that is €100 in the black. For your selling decision, you will obviously regard the buying price of €1,000 as the reference point. The coded decision problem thus describes a choice between a sure gain of €100 and a risky lottery that offers a gain of €200 or €0 with equal probability. As your PT value function is concave for gains, you will act as risk averse and choose the sure gain of €100; you thus sell the winner stock.

Had you instead purchased the stock at a price of €1,200, it would now constitute a loser stock. Your selling decision can be described as a choice between a sure loss of €100 and a risky lottery that offers a gain of €200 or €0 with equal probability. The convex shape of the PT value function in the domain of losses lets you behave as risk seeking so that you will not realize the loss but rather hold on to the stock and take the risk. These connections are illustrated in Figure 13.11. The investor behavior in the disposition effect thus exactly corresponds to the predictions of prospect theory and can not only be observed with real portfolio data but also in very controlled experimental studies that can exclude other explanations (see Weber and Camerer 1998). More recent studies are a little skeptical of this very simple PT-based explanation (see Barberis and Xiong 2009) but we do not want to address these subtle deliberations because of the intuitive appeal of the effect.

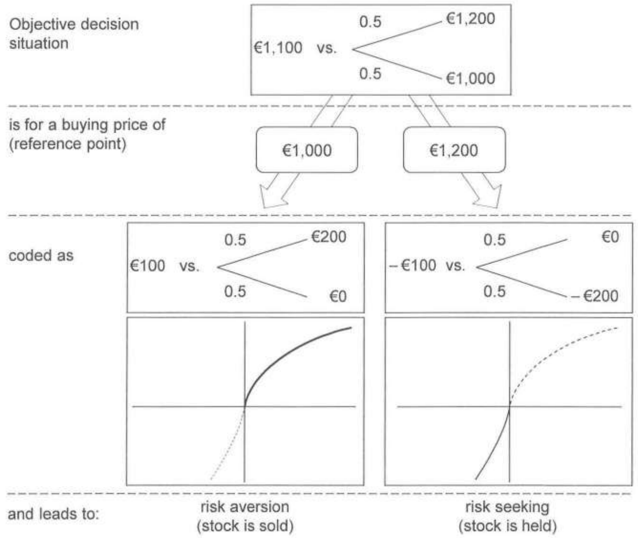


Figure 13-12: Explanation of the disposition effect by prospect theory

The equity premium puzzle and myopic loss aversion

In general, equities have much more volatile returns than do government bonds (considering countries where the repayment can be rated as reliable). To compensate the investor for bearing the additional risk, the long-run return of equities correspondingly is higher (the equity premium). For long parts of the past century, overall average stock returns were about 8% higher than the returns on government bonds. At first, this was regarded as an adequate *risk premium* for stocks, i.e. it was assumed that for the average risk-averse investor stocks are equally attractive as bonds for precisely this risk premium. Mehra and Prescott (1985) analyzed the level of risk aversion implied by this risk premium and found that - under the classic assumptions of expected utility theory - investors must have an unrealistically high risk aversion: an investor needed to be indifferent between a 50/50 lottery offering a wealth level of €50,000 or €100,000 and a certain wealth of €51,209.

Since then, economists have tried to find a rational explanation for the high equity premium but have not really been successful. Benartzi and Thaler (1995) proposed a behavioral explanation based on prospect theory. The core idea of the explanation is the assumption that investors can have long investment horizons (where the volatile stock returns can offset each other) yet their evaluation period is much shorter. Such myopia can be interpreted as a special intertemporal type of mental accounting (often referred to as "narrow bracketing" in the intertemporal context; see also Section 13.2.5). The investor evaluates the returns of investments in mental accounts without taking note of the fact that, from a superordinate perspective (considering the entire investment horizon), other evaluations would result; due to this myopia, his/her loss aversion is especially pronounced. While for long investment horizons, stocks have hardly any risk of loss in practice, it is quite probable for a stock investment to be faced with a loss on a yearly basis. Benartzi and Thaler now argue that a myopically loss-averse investor negatively perceives this potential of loss and demands substantial compensation by means of a higher expected return.

To supplement their theory of myopic loss aversion by data, Benartzi and Thaler calculate the CPT evaluations for different evaluation periods based on historic returns for stocks and bonds. The shorter the evaluation period, i.e. the more myopic the evaluation of investments, the less attractive the stocks are to the investor as losses become increasingly probable. We want to illustrate this connection with the help of an example.

A risky investment can yield a loss of €100 or a gain of €200 with equal probability; it hence provides a gain of €50 in expectation. The riskless asset offers a sure gain of €10. Regarding the decision rule of the investor, we want to focus solely on loss aversion and assume that no probability weighting is carried out and diminishing value sensitivity also does not exist. The value function of the investor has the simple form:

$$v(x_i) = \begin{cases} x_i & \text{for } x_i \geq 0 \\ 2.25 \cdot x_i & \text{for } x_i < 0 \end{cases} \tag{13.13}$$

The investor wants to invest for two periods and can thus receive either two independent realizations of the risky lottery or twice the sure gain. If the investor is not myopic, he recognizes that the risky investment only has a 25% chance of a loss of €200 (if things turn out badly in both periods), while he receives a gain of €400 with 25% probability and a gain of €100 with 50% probability (see Figure 13-12). Using the value function of (13.13) he assigns a value of 37.5 to the lottery (–€200, 0.25; €100, 0.5; €400, 0.25) and thus finds it more attractive than the twofold riskless investment, evaluated at 20. Nevertheless, when the investor acts myopically he will compare a single realization of the risky investment to the (single) riskless asset. Because of his strong loss aversion, this turns out to be very unattractive to him; he calculates an evaluation of –12.5. The myopic investor thus prefers the riskless investment.

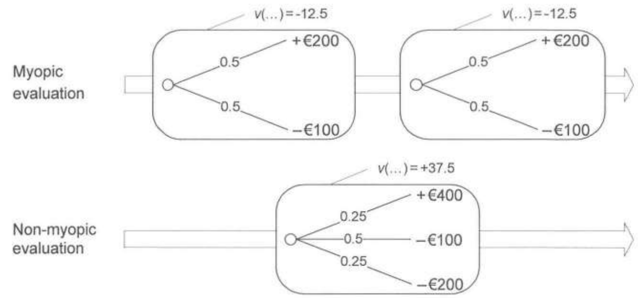


Figure 13-13: Evaluation effects of myopic loss aversion

Benartzi and Thaler now investigate how long the evaluation period must be for stocks and bonds to be historically equally attractive to a CPT decision maker. They find it to hold for an evaluation period of 12 months, a number that seems very plausible when keeping in mind that historic stock returns are generally presented on a yearly basis and that investors receive information on the performance of their assets on their year-end securities account statements.

Later studies have also found additional evidence for the concept of myopic loss aversion. In experiments where the myopia of participants was for instance manipulated by more frequent investment feedback (Gneezy and Potters 1997), a systematic relationship between myopia and the willingness to take risks could be observed; even the more detailed predictions of a myopic prospect theory could be validated experimentally (Langer and Weber 2005).

13.3.4 Further theories – *disappointment* and *regret*

At the start of this section, we would like to review the different concepts of evaluation that we have discussed so far. Based on the simplest evaluation using the expected value

$$\text{EW}(a) = \sum_{i=1}^{n} p_i \cdot a_i \,, \tag{13.14}$$

which directly weights the consequences with the associated probabilities, the formula of expected utility theory

$$\text{EU}(a) = \sum_{i=1}^{n} p_i \cdot u(a_i) \tag{13.15}$$

transforms the consequences yet maintains the additive-separable form so that for every consequence, an independent contribution to the overall evaluation is calculated and then added up.

Also, the original Prospect Theory (OPT) of Kahneman and Tversky (1979) that we have not discussed in detail preserved this general form:

$$\mathrm{OPT}(a) = \sum_{i=1}^{n} w(p_i) \cdot v(a_i)\,, \tag{13.16}$$

even if, because of probability weighting, no true expected values were considered any longer.

Additive separability was no longer given for RDEU theories and CPT due to the cumulative probability weighting

$$\mathrm{RDEU}(a) = \sum_{i=1}^{n} w(p_1, p_2, \ldots, p_i) \cdot u(a_i)\,. \tag{13.17}$$

In this case, the weight of an evaluated consequence a_i is not only determined by its probability of occurrence p_i but the probabilities $p_1, p_2, \ldots, p_{i-1}$ also have an influence. CPT (and OPT) additionally coded the consequences, i.e. they were defined with respect to a reference point.

The disappointment theories we are about to review model other emotional factors and can easily be integrated into this spectrum of evaluation models. They are also based on the idea that the decision maker uses a reference point during the evaluation of consequences. In contrast to the assumptions of prospect theory, this reference point is not exogenously given but implicitly contained in the lottery to be evaluated. Put in a general formula, it can be denoted as:

$$D(a) = \sum_{i=1}^{n} p_i \cdot u(a_i, a) \tag{13.18}$$

For the disappointment theories, additive separability during evaluation is also no longer given; when calculating every addend, information on the other consequences and on the probabilities of the other states is required.

Disappointment theories

The main idea of disappointment theories can be illustrated with the help of a simple decision problem: let us assume a decision maker were to participate in the two lotteries (€10,000, 0.99; €0, 0.01) and (–€10,000, 0.99; €0, 0.01). After the results of the lotteries have been determined, the decision maker is informed that he/she did not win in the former case and did not lose in the latter case.

Many decision makers will perceive the result of €0 very differently across both lotteries: for the first lottery, they will certainly be *disappointed*, and most decision makers will feel *elated* about the result of the second lottery. The context of the realized consequence of €0 seems to impact the evaluation of the consequence. Utility theory does not allow for considering the context when evaluating consequences. As pointed out by formula (13.15), the utility of the consequence "€0" is necessarily independent of the other consequences of an alternative. While for the rank-dependent theories, the rank of a consequence determines the strength

of probability transformation, disappointment theories explicitly model the context effect of a value of a consequence (for disappointment theories see Bell 1985 and Loomes and Sugden 1986). For the connection between disappointment theories and RDEU theories see also Delquié and Cillo (2006).

When decision makers anticipate possible disappointment and elation effects in their decision, these effects can have an influence on the decision itself; for example, they may forego the chance of a higher gain to avoid possible disappointment.

The theories assume the existence of a measurable value function v. For a risky alternative, decision makers are disappointed or elated if the result does not meet the expectations or exceeds the expectations, respectively. The proposal was to define the expected value $\mathrm{EV}(v(a))$ of the distribution of consequences as the expectation of a decision maker towards (the outcome of) a lottery. We want to denote this expectation that defines the benchmark of whether a decision maker is disappointed or elated by an outcome, by v^*. The utility of a consequence $u(a_i)$ is defined as the sum of the value of the consequence plus the associated disappointment or elation:

$$u(a_i) = v(a_i) + D(v(a_i) - v^*) \tag{13.19}$$

$D(v(a_i) - v^*)$ is a function mapping the associated disappointment or elation. The disappointment utility (DE utility) of an alternative is defined by:

$$\mathrm{DE}(a) = \sum_{i=1}^{n} p_i(v(a_i) + D(v(a_i) - v^*)). \tag{13.20}$$

If $D(\cdot)$ is linear, the terms for disappointment and elation offset each other exactly. The disappointment utility reduces to the expected value of the consequences evaluated by the value function. The preference generated by DE is transitive; it conforms to the condition of stochastic dominance, if $D(\cdot)$ is an increasing function with a slope below 1. It is further assumed that $D(\cdot)$ is convex for positive arguments (elation) and concave for negative arguments (disappointment).

An axiomatic foundation for disappointment theory can be found in Gul (1991). A weakening of the independence axiom is again pivotal for this approach. Simplified, it can be said that – in contrast to utility theory – the independence axiom must not hold in general. It must only be applied to those alternatives where the application does not generate disappointment or elation effects. Disappointment theory is also capable of explaining Allais-paradox behavior, an argument that is supported by the shape of the indifference curve in the three-outcome-diagram in Figure 13-13.

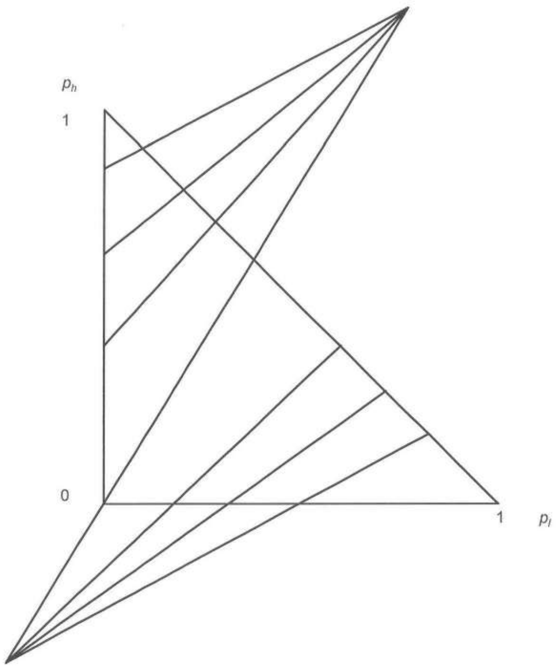


Figure 13-14: Utility indifference curves for disappointment theory

Regret theory

In regret theories, an emotional factor is also accounted for. At first glance, regret theories seem to be very similar to disappointment theories. However, if you attempt to classify regret theories within the concept of evaluation of (13.14), you will soon notice that their conceptual approach extends far beyond this. The core idea of regret theories is to no longer evaluate the utility of an alternative in isolation but to evaluate pairs of alternatives jointly (see Bell 1982, Loomes and Sugden 1982 and Fishburn 1984). When comparing two alternatives, a decision maker anticipates the "regret" that he might feel when experiencing the decision outcome. This pairwise consideration allows for modeling of intransitive preferences: a can be preferred to b (when a and b are jointly evaluated), b to c, and c to a.

An example (see Table 13-5) is intended to clarify the regret theory's way of thinking (see, for example, Paterson and Diekmann 1988). At the same time, the example illustrates the additional potential of regret theories of capturing diverse preferences.

In the example, the consequences of two alternatives a and b are presented that are generated by the roll of a fair die labeled with the numbers 1 to 6. As all states in the example occur with the same probability of p = 1/6, the consequences and probability distributions of a and b are identical. All previously presented preference theories predict that every decision maker is indifferent between the two alternatives. As you might recognize from your own preferences, one can put forward good reasons for a preference $a \succ b$ or $a \prec b$. In an empirical investigation by Paterson and Diekmann (1988), 84% of the participants preferred alternative a.

Table 13-5: Example of the relevance of regret theories

State (number on die)	1	2	3	4	5	6
a	20	30	40	50	60	10
b	10	20	30	40	50	60

To be able to model a preference between a and b, a theory must define the consequences with respect to the states. It can then be determined what consequences (of the different alternatives) will occur in which state. Building on this, it becomes possible to evaluate the consequences of an alternative with reference to the consequences of another alternative. This evaluation of consequences depending on the consequences of a second alternative is just the characteristic feature of regret theories and enables these theories to model a preference between a and b in our example. Formally, regret theory (ER: expected regret) can be denoted as:

$$\mathrm{ER}(a,b)=\sum_{i=1}^{n} p(s_i)\,(v(a_i)+R(v(a_i)-v(b_i))), \tag{13.21}$$

where a_i and b_i are the consequences of the alternatives a and b in case state s_i occurs. In the form

$$\mathrm{ER}(a,b)=\sum_{i=1}^{n} p(s_i)\cdot v(a_i)+\sum_{i=1}^{n} p(s_i)\cdot R(v(a_i)-v(b_i)) \tag{13.22}$$

it becomes even more evident that in regret theory, the expected value of a lottery is adjusted for the expected regret.

R is the strictly monotonically increasing regret function that quantifies the regret (or the elation) regarding the occurrence of a_i instead of b_i in case state i occurs. An alternative a is preferred to an alternative b if and only if $\mathrm{ER}(a, b) > \mathrm{ER}(b, a)$.

Regret theory models a common feeling during decision making. To explain the theory is some more detail, let us have a look at the choice between a sure alternative a = €3,000 and the lottery b = (€0, s_1; €4,000, s_2) with $p(s_1) = 0.2$ and $p(s_2) = 0.8$; a choice that you are already familiar with from the Allais paradox. For linear value functions, it holds that:

$$\mathrm{ER}(a,b)=3{,}000+0.2\,R(3{,}000-0)+0.8\,R(3{,}000-4{,}000),$$

$$\mathrm{ER}(b,a)=3{,}200+0.2\,R(0-3{,}000)+0.8\,R(4{,}000-3{,}000).$$

You immediately see that the regret function can lead to a preference for alternative a. The term "$R(0-3{,}000)$" in the expression $\mathrm{ER}(b, a)$, i.e. the regret that occurs when you receive nothing if you could have received €3,000 when choosing the other lottery, can be sufficient to prompt people to select a despite the higher expected value of b. Zeelenberg and Pieters (2004), for example, shows that regret aversion can also be documented in decision situations of everyday life.

You can also recognize the strong conceptual differences between regret theory and the other theories by the fact that we cannot present you with a probability triangle of utility indifference curves for regret theory. To be able to draw these curves, every lottery needs to be assigned a clearly defined utility. In the case of regret theory this is not the case because the utility of a lottery is influenced by the decision environment (precisely, by the set of the other available alternatives).

13.3.5 Current developments in the area of descriptive decision theory

Over the last years, researchers have continued their work on descriptive preference theories. At this point, we want to present two promising developments: third-generation prospect theory (PT^3) and the system1/system2 concept.

The *third-generation prospect theory* by Schmidt et al. (2008) supplements reference-dependent evaluation and probability weighting by "uncertain" reference points. Similar to regret theory, PT^3 does not evaluate consequences in isolation but relative to a reference act. This reference act can differ depending on the framing of the decision problem, i.e. it can be "uncertain." If the reference act is always equal to the status quo, PT^3 simplifies to CPT. With the help of this extension, PT^3 is able to explain standard patterns of various preference reversals while maintaining the predictive ability of CPT.

The *system1/system2 concept* starts from a more basic psychological level, but is no less interesting for it. The term system1/system2 was coined for specific types of dual-process models (Kahneman 2003). System 1 refers to intuitive decision-making behavior; system 2 refers to logical reasoning based on deliberate considerations. In its intuitive mode of operation, system 1 is fast and automated and works effortlessly and associatively. However, it also succumbs to many biases, for instance the availability bias. System 2 is a slower process that is deliberate and rule-based but also very strenuous. Both systems do not necessarily have to act separately but can also work together. System 1 provides a fast intuitive answer, and, if necessary, system 2 interferes to correct the mistakes of system 1 via a controlled process. System 2 knows about the typical mistakes of system 1 and takes care to correct these. However, if it is not successful at doing so, systematically biased assessments and decisions result. The phenomenon of anchoring and adjustment, for instance, can be interpreted from a system1/system2 perspective as follows. System 1 defines the starting point (anchor) and is easily influenced by external inputs. System 2 ideally recognizes the error and adjusts accordingly. If this adjustment is too small, the judgment remains biased towards the anchor.

A neat example where in most cases system 2 does not take a corrective action is presented by Kahneman in his 2002 Nobel prize lecture: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" For a moment (due to the fast and intuitive system 1) nearly everyone thinks that the answer must of course be 10 cents. System 2 recognizes the false conclusion and corrects the answer – mentally – to 5 cents. It becomes clear that depending on the type of decision situation (and on the decision maker), system 1 and system 2 cooperate differently.

13.4 Conclusion

If one was to summarize this chapter, it would soon become clear that decision behavior cannot be explained by traditional utility theory. Accounts of paradoxes and other typical behavioral patterns that cannot be represented by expected utility theory are accumulating quickly. You do not have to be concerned about this as expected utility theory is meant to support rational behavior and does not raise a descriptive claim.

However, for different reasons this chapter's insights matter to those readers who are interested in rational decision making. On the one hand, they strikingly exemplify the importance of a structured decision-making process that is particularly concerned with rationality. Decision making that is based on intuition will often lead to systematic biases and serious mistakes. Detailed knowledge of typical mistakes in decision making can raise awareness in the sense of system1/system2 thinking of when it is particularly meaningful that the rule-based system2 is activated to correct the presumed spontaneous misjudgment of system1.

Beyond that, the descriptive insights are essential for the development of better methods for decision support. You have become acquainted with a typical example in Chapter 9; all the methods for the determination of utility functions were based on the interpretation of choices and certainty equivalents. In doing so, it was implicitly assumed that the respondents include unbiased probabilities in their decision calculus. The descriptive insight that probabilities are perceived in a systematically biased way lets standard methods appear problematic. A remedy may be obtained through improved methods that are immune to biases (like the trade-off method for utility functions in Section 9.4.6). Alternatively, the results of the elicitation methods can explicitly be adjusted for these errors. Such an approach is suggested by Bleichrodt et al. (2001). It is presumed that the decision maker uses a probability weighting function when stating her certainty equivalents that is equal to the function proposed by Tversky and Kahneman (1992) including the respective (median) parameters. It is for instance hence assumed that the decision maker perceives the probability of her preferred consequence in a 50/50 lottery as being only 42.1% when stating the certainty equivalent. This is a convincing approach for mean values (groups of decision makers). Based on a set of experimental data, Bleichrodt et al. (2001) showed that the observed biases can be almost completely eliminated on the group level using the proposed corrections.

Finally, the descriptive insights are of particular importance for the concept of liberal paternalism (Sunstein and Thaler 2003). The idea of liberal paternalism is to shape the decision situation (accounting for the descriptive insights) in such a way that it is made simple for the decision maker to choose the best alternative. For an example, let's regard the tendency to eat too much of food that is too fattening when going out for a meal in a restaurant (as you know from reading this book, this is a typical problem of hyperbolic discounting: today's pleasure is given too much weight compared with the negative consequences in the far future). From the paternalistic point of view, the diners should be influenced towards having healthier food in order to maximize their long-term utility. This contradicts the common model of mature citizens who will not be patronized with respect to their decisions. If descriptive decision research was to find that it is primarily courses from the first page of the menu that are ordered in restaurants, the eating behavior of the customers could be influenced without restricting their leeway: from the paternalistic perspective, there should be healthy food on the first page of the menu. With similar "tricks" and "nudges" (Thaler and Sunstein 2009), people can be made to increase their retirement savings or to avoid unnecessary risks when driving a car.

Questions and exercises

13.1

At the time of observation a stock's price is 250; the investor originally bought this stock at 100. Regarding the future development of the stock price, it is known that it will either rise to 350 with a probability of 0.6 or fall to 150 with a probability of 0.4. The investor evaluates risky alternatives according to prospect theory. His value function for gains and losses relative to the given reference point is

$$\begin{array}{c} x^{0.88} \text{ for } x \geq 0 \\ -2.25(-x)^{0.88} \text{ for } x \leq 0. \end{array}$$

For reasons of simplicity, it is assumed that the investor does not transform probabilities (i.e. it holds that $\pi(p) = p$ for all p).

Show that the optimal decision according to prospect theory considering the two alternatives "hold stock" and "sell stock" depends on whether the reference point is given by the historical or the current price of the stock.

13.2

A criminal has to choose the place to which he wants to break in at night. The three potential targets with their corresponding goods are given in the following table:

Object	Type of goods	Amount of goods
Newsstand	certain	121
Diner	risky	evenly distributed over [50; 250]
Jeweler	risky	0 (prob. = 0.75) or 900 (prob. = 0.25)

The criminal evaluates alternatives according to prospect theory. His value function for gains and losses relative to the given reference point is

$$\begin{array}{c} x \text{ for } x \geq 0 \\ k \cdot x \text{ for } x \leq 0 \ (k \geq 1). \end{array}$$

For reasons of simplicity, it is assumed that the criminal does not transform probabilities (i.e. it holds that $\pi(p) = $ p for all p).

His reference point is 100 as given by the (certain) income from his time robbing slot machines.

(a) Which alternative is he going to choose if his loss-aversion coefficient is equal to $k = 2$?
(b) How does his decision change for $k = 6$?

However, there is one thing unknown to the criminal: he faces a security guard as his opponent. However, the security man can only guard one object per night. If the criminal selects the guarded object he has to leave without goods. The security guard maximizes his expected utility. His utility function m for goods (or monetary equivalents) that could not be stolen due to his service is given by $u(m) = m^{0.5}$. The security guard does not know the exact value of the criminal's loss-aversion coefficient k; he assumes the values $k = 2$ and $k = 6$ to be evenly likely.

(c) Which object will the security guard choose to protect?
(d) What does the probability distribution of the criminal's loss-aversion coefficient k have to look like so that the security guard can use a coin toss to decide which object he should protect later at night?
(e) To what extent would the security guard's approach become more complicated if the (certain) goods from robbing the newsstand were only 118.75 instead of 121?
(f) The criminal is much cleverer than assumed before; he now knows about the security guard's existence and is even informed about the guard's decision procedure (and hence knows his strategy). However, the security guard does not know that the criminal possesses this information. Show that it holds for $k = 6$ that the strategy chosen in (b) is no longer optimal.

13.3

Decision maker A and decision maker B evaluate risky alternatives according to a rank-dependent utility function.

While both show identical utility functions for money $u_A(x) = u_B(x) = x$, they apply different probability transformation functions: for decision maker A it holds that $g_A(p) = p^{0.5}$ and for decision maker B it holds that $g_B(p) = p$.

The uncertainty is described by three possible states of the world that occur with the probabilities as given by the following table. Furthermore, the table shows the state-dependent wealth levels of the decision makers. For instance, decision maker B's wealth level amounts to €4 if state s_3 is realized.

State	s_1	s_2	s_3
$p(s_i)$	0.25	0.51	0.24
Decision maker A	0	8	3
Decision maker B	6	2	4

Decision maker B intuitively senses that the given allocation of conditional wealth claims offers room for barter. The decision makers agree upon the following negotiating structure: decision maker B is allowed to suggest one exchange possibility (e.g. "I offer €3 in state s_3 and want to receive €2 in state s_2"), which decision maker A can either reject or accept (whereas we assume that in the case of A's indifference between the current solution and the new suggestion he accepts B's proposal).

(a) Determine the decision makers' utility for the initial situation before any barter.
(b) Why does the initial situation allow for exchange possibilities benefiting both sides in the first place?
(c) What offer will decision maker B submit to decision maker A if he wants to maximize his utility? What will decision maker A's response look like? Determine the utility for both decision makers resulting from the allocation after the barter being realized.
(d) Give a verbal explanation for how far the distribution of "*gains from trade*" (in the sense of utility change due to barter) is determined by the choice of negotiating structure.

13.4

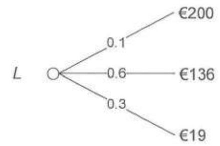
Peter decides according to cumulative prospect theory (CPT) and has the value function:

$$v(x)=\begin{cases} x^{\alpha} & \text{for } x\geq 0 \\ -k\cdot(-x)^{\alpha} & \text{for } x<0 \end{cases}$$

with $k>1$ and $0<\alpha<1$.

(a) Is CPT a descriptive or prescriptive decision theory? Explain the main difference between the two kinds of theories. Is it possible for a good prescriptive theory to be a good descriptive theory at the same time?
(b) What properties of prospect theory are represented by the parameters α and k?

(c) Let us assume that Peter's prospect theory parameters are $\alpha = 0{,}5$ and $k = 2$. He evaluates the following lottery



and codes the consequences relative to his reference point (€100). Would he accept the lottery if he did not apply probability weighting?
(d) Now suppose that Peter would apply probability weighting using the same transformation function for gains and losses: $g(=g^{+}=g^{-})$. The following table summarizes some values of the function g:

prob.	10%	20%	30%	50%	60%	70%	80%	90%
g(prob.)	15%	18%	20%	38%	45%	55%	65%	80%

Would he accept the lottery now?

13.5

A typical value function in prospect theory implies the following kind of strategy: "Do not pack all presents into one box." More formally, one could say that the recipient of the presents would be more delighted about the presents (and their respective positive consequences) if he evaluated them *separately* instead of evaluating them in a combined *package*.

(a) Which property of the value function in prospect theory causes this effect and why?
(b) Draw an exemplary value function assumed by prospect theory that indicates all the relevant theoretical properties and state their names.
(c) What would a corresponding recommendation regarding single packaging vs. one combined package look like when it comes to negative instead of positive consequences?
(d) When comparing the evaluation of a package combining a gain and a loss with the separate evaluation of the two components, the magnitude of the gain and the loss plays an important role. For which of these combinations might the separate evaluation be more attractive than the evaluation of the combined package?
  (1) [large gain + small loss] or
  (2) [small gain + large loss]
(e) Why can such a "Do not pack all presents into one box" effect not be relevant for expected utility maximizers?

13.6

MonaLisA (MLA) is a myopic loss-averse decision maker who is asked to participate in a sequence of (stochastically independent) lotteries of the form (+€200, 0.5; –€100, 0.5). Monalisa does not perform probability weighting; her value function is piecewise linear (i.e. it holds that $\alpha = \beta = 1$). However, she is strongly loss averse with $k = 3$. Monalisa is not extremely myopic, which means she does not evaluate each single lottery but rather considers combinations of a few lotteries, but not the whole sequence of lotteries as she actually should. How myopic is Monalisa maximally allowed to be if she was still to accept the whole sequence of lotteries (i.e. at least how many lotteries does she have to combine for the resulting distribution to be evaluated positively by her despite her loss aversion)?