A variety of approaches exist to formalise problems of action planning under uncertainty, each of them emphasising other aspects of this type of problem. As motivated in Chapter 1, in this thesis we choose to employ (Bayesian) decision theory at the fundamental level of comparing alternative actions in situations of choice. The basics of decision theory are discussed in this chapter. Decision theory can be regarded as synthesising (Bayesian) probability theory and utility theory. Here, probability theory serves as a framework for reasoning with uncertainty, whereas utility theory provides the guidelines for rational choice under uncertainty.

This chapter is structured as follows. We first discuss probability theory in Section 3.1, and continue with utility theory in Section 3.2. The synthesis of both theories is given in Section 3.3, which also provides an overview of how different types of decision problem are analysed with this type of reasoning. We conclude with a short discussion in Section 3.4. The chapter also serves to introduce a number of formal notations that will be used in subsequent chapters. Throughout, the various notions that are introduced will be illustrated with examples from the domain of paediatric cardiology as described in the previous chapter.

### 3.1 Probability theory

In this section, we review the main concepts from probability theory. Our review will be concise, and is not intended to be exhaustive: it highlights the aspects of
probability theory that are crucial in decision-theoretic reasoning; for a thorough introduction to the theory, we refer to (Shiryayev, 1984; Grimmett and Stirzaker, 1992). We start introducing the language we will employ to describe elements from a domain of interest; this language, a Boolean algebra, consists of Boolean expressions over value assignments to a given set of variables (Birkhoff and MacLane, 1977). Let \( \text{dom}(w) \) denote the domain of a given variable \( w \), i.e. the set of possible values that the variable may take.

**Definition 3.1 (Boolean algebra)** Let \( W \) be a set of variables. The Boolean algebra spanned by \( W \), denoted by \( \beta(W) \), comprises all expressions built up from value assignments to elements from \( W \), the constants \( T \) (true) and \( \bot \) (false), the binary operators \( \land \) (conjunction) and \( \lor \) (disjunction), and the unary operator \( \lnot \) (negation).

In the Boolean algebra \( \beta(W) \), value assignments to elements from \( W \), i.e. expressions of the form \( w = v \), where \( w \in W \) and \( v \in \text{dom}(w) \), act as Boolean variables. We use \( \varphi \equiv \psi \) to indicate that the expressions \( \varphi, \psi \in \beta(W) \) are equivalent under the usual axioms. Furthermore, we will write \( \varphi \vdash \psi \) when \( \varphi \equiv \varphi \land \psi \), or equally \( \psi \equiv \varphi \lor \psi \). The constants \( \bot \) and \( T \) now denote the universal upper and lower bounds of the distributive and complemented lattice on \( \beta(W) \) induced by the relation \( \vdash \).

Within our theory, a prominent part is played by conjunctions of value assignments to sets of variables; we will refer to such conjunctions as configurations.

**Definition 3.2 (Configuration)** Let \( W \) be a set of variables. A conjunction

\[
\text{c}_W = \bigwedge_{w \in W} w = v
\]

(3.1)

of value assignments to the variables from \( W \) is called a configuration of \( W \). The set of all configurations of \( W \) is called the universe of \( W \), notation \( \Omega_W \).

Note that \( \Omega_W \subseteq \beta(W) \). Furthermore, there is but one configuration of the empty set \( \emptyset \), and this is the empty conjunction \( T \). We will usually write \( \text{c}_W \) to denote a configuration of \( W \) (i.e. \( \text{c}_W \in \Omega_W \)); for a singleton set \( \{w\} \), we will simply write \( \text{c}_w \) and \( \Omega_w \) instead of \( \text{c}_{\{w\}} \) and \( \Omega_{\{w\}} \), respectively. So,

\[
\Omega_w = \{ w = v \mid v \in \text{dom}(w) \}
\]

(3.2)

for each variable \( w \).

**Notation 3.3** Let \( W = \{w_1, \ldots, w_n\} \) be a set of variables. We use

\[
\text{dom}(W) = \text{dom}(w_1) \times \cdots \times \text{dom}(w_n)
\]

(3.3)

to denote the set of possible values of \( W \), and \( W = V \), where \( V \in \text{dom}(W) \), \( V = \{v_1, \ldots, v_n\} \), as an alternative notation for the configuration

\[
\text{c}_W \equiv w_1 \land \cdots \land w_n = v_n.
\]

(3.4)
### 3.1 Probability theory

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSD</td>
<td>VSD size</td>
<td>null, small, moderate, large</td>
</tr>
<tr>
<td>shunt</td>
<td>shunt size</td>
<td>none, small, moderate, large, reversed</td>
</tr>
<tr>
<td>resis</td>
<td>pulmonary vascular resistance</td>
<td>normal, increased, high, very high</td>
</tr>
<tr>
<td>hfail</td>
<td>heart failure</td>
<td>absent, mild, moderate, severe</td>
</tr>
<tr>
<td>pmhyp</td>
<td>pulmonary hypertension</td>
<td>absent, mild, moderate, severe</td>
</tr>
<tr>
<td>pmart</td>
<td>pulmonary arteriopathy</td>
<td>false, true</td>
</tr>
<tr>
<td>closure</td>
<td>spontaneous closure</td>
<td>false, true</td>
</tr>
<tr>
<td>death</td>
<td>death</td>
<td>false, true</td>
</tr>
</tbody>
</table>

**Table 3.1:** Example variables for the VSD domain.

We will regard the expression $W = V$ as an element of the Boolean algebra $\beta(W)$.

**Notation 3.4** Let $W$ be a set of variables. The expression $C_W = \bigwedge_{w \in W} w$ is called the configuration template of $W$. From the configuration template of $W$, any configuration $c_W \in \Omega_W$ can be obtained by substituting the variables with appropriate value assignments.

We will generally use configuration templates within formulas. The templates then provide for treating the formulas as schemata from which multiple instantiations can be obtained by filling in values for the variables in the templates.

Given two sets $Y$ and $Z$ of variables, we say that their configurations $c_Y$ and $c_Z$ are compatible when $c_Y \land c_Z \neq \bot$. This holds when either $Y$ and $Z$ are disjoint, or $c_Y$ and $c_Z$ assign the same values to variables in $Y \cap Z$. Otherwise, these configurations are called incompatible.

**Example 3.5** Consider the set of variables listed in Table 3.1, representing concepts from the domain of paediatric cardiology; each of the variables describes a part of the clinical state of VSD patients. Example configurations from the Boolean algebra of propositions spanned by this set are

\[
\varphi_1 : \text{VSD} = \text{small} \land \text{shunt} = \text{small}, \\
\varphi_2 : \text{shunt} = \text{small} \land \text{hfail} = \text{mild}, \text{ and} \\
\varphi_3 : \text{shunt} = \text{large} \land \text{hfail} = \text{severe}.
\]

We have that $\varphi_1 \in \Omega_{\{\text{size,shunt}\}}$, and $\varphi_2, \varphi_3 \in \Omega_{\{\text{shunt,hfail}\}}$. Furthermore, $\varphi_1$ and $\varphi_2$ are compatible, whereas neither $\varphi_1$ and $\varphi_3$, nor $\varphi_2$ and $\varphi_3$ are.

We will now consider a set $X$ of random variables, and define a joint probability distribution on $X$ as a function on the Boolean algebra of propositions spanned by $X$. 
Definition 3.6 (Probability distribution) Let $X$ be a set of random variables, and let $P : \beta(X) \to [0, 1]$ be a function such that

1. $P(\top) = 1$,
2. $P(\varphi) = 0$ when $\varphi \equiv \bot$, and
3. if $\varphi, \psi \in \beta(X)$ such that $\varphi \land \psi \equiv \bot$, then $P(\varphi \lor \psi) = P(\varphi) + P(\psi)$.

Then, $P$ is called a joint probability distribution on $X$. For each $\varphi \in \beta(X)$, the function value $P(\varphi)$ is termed the probability of $\varphi$.

Note that we associate probabilities with Boolean expressions instead of with sets, which is common in textbooks on probability theory. Both notions of probability are however equally expressive, (de Finetti, 1970). We say that a joint probability distribution $P$ on $X$ is degenerate when $P(c_X) \in \{0, 1\}$ for all $c_X \in \Omega_X$; otherwise, it is non-degenerate. A non-degenerate distribution $P$ is called strictly positive when $P(c_X) > 0$ for all $c_X \in \Omega_X$; it is uniform when all configurations of $X$ have equal probability, i.e. when $P(c_X) = 1/|\Omega_X|$ for all $c_X \in \Omega_X$.

Definition 3.7 (Conditional probability) Let $X$ be a set of random variables, and let $P$ be a joint probability distribution on $X$. For each $\varphi, \psi \in \beta(X)$ such that $P(\psi) > 0$, the conditional probability of $\varphi$ given $\psi$ is defined as

$$P(\varphi \mid \psi) = \frac{P(\varphi \land \psi)}{P(\psi)}.$$  \hfill (3.5)

The conditional probability $P(\varphi \mid \psi)$ expresses the amount of certainty concerning the truth of proposition $\varphi$, given that the information $\psi$ is known with certainty. We state without proof that for a given $\psi \in \beta(X)$ with $P(\psi) > 0$, the conditional probabilities $P(\varphi \mid \psi)$ for all $\varphi \in \beta(X)$ once more constitute a joint probability distribution on $X$. This distribution is called the conditional probability distribution given $\psi$, and we also denote it by $P^\psi$. Conditional probability distributions are also referred to as posterior distributions, as they describe the respective probabilities posterior to the incorporation of a piece of information; the original (unconditional) probability distribution is then called the prior distribution.

The notion of conditional independence (Dawid, 1979) allows for qualification of the relations between variables in a joint probability distribution; it underlies the graphical representation of probability distributions which is discussed in Chapter 4.

Definition 3.8 (Conditional independence) Let $X$ be a set of random variables, let $Y_1, Y_2, Z \subseteq X$, and let $P$ be a joint probability distribution on $X$. Then, the set
3.1 Probability theory

$Y_1$ is said to be conditionally independent of the set $Y_2$ given the set $Z$ under the distribution $P$, notation $Y_1 \perp_{P} Y_2 \mid Z$, if

$$P(C_{Y_1} \mid C_{Y_2} \land C_{Z}) = P(C_{Y_1} \mid C_{Z}).$$

(3.6)

If $Z = \emptyset$, we say that $Y_1$ is marginally independent, or simply independent, of $Y_2$, and write $Y_1 \perp_{P} Y_2$.

Intuitively, $Y_1 \perp_{P} Y_2 \mid Z$ means that if $Z$ is known, then determining $Y_2$ provides no further knowledge about $Y_1$. Recall that $C_{Y_1}$, $C_{Y_2}$, and $C_{Z}$ denote the configuration templates of $Y_1$, $Y_2$, and $Z$, respectively, representing all possible configurations of these sets. Note that conditional independence of $Y_1$ from $Y_2$ given $Z$ is equivalent with

$$P(C_{Y_1} \land C_{Y_2} \mid C_{Z}) = P(C_{Y_1} \mid C_{Z}) \cdot P(C_{Y_2} \mid C_{Z}).$$

(3.7)

Therefore, conditional independence is symmetrical in $Y_1$ and $Y_2$.

We now recapitulate some well-known propositions from probability theory that allow to make inferences from a given probability distribution. We assume that $X$ is a set of random variables, $P$ is a strictly-positive joint probability distribution on $X$, and that $Y$ and $Z$ are arbitrary subsets of $X$. Note that a joint probability distribution $P'$ on $\beta(Y)$ can be derived from the probabilities $P(c_Y)$ for all configurations of $Y$; $P'$ is then called the marginal distribution on $Y$.

**Proposition 3.9 (Chain rule)**

$$P(C_X) = P(C_{X \setminus Y \cup Z} \mid C_{Y \cup Z}) \cdot P(C_Y \mid C_Z) \cdot P(C_Z).$$

(3.8)

**Proof.**

$$P(C_{X \setminus Y \cup Z} \mid C_{Y \cup Z}) \cdot P(C_Y \mid C_Z) \cdot P(C_Z) =
\frac{P(C_X \setminus Y \cup Z \land C_{Y \cup Z})}{P(C_{Y \cup Z})} \cdot \frac{P(C_Y \land C_Z)}{P(C_Z)} \cdot P(C_Z) =
\frac{P(C_X)}{P(C_{Y \cup Z})} \cdot \frac{P(C_Y \cup Z)}{P(C_Z)} \cdot P(C_Z) = P(C_X)$$

(3.9)

\[ \square \]

**Proposition 3.10 (Marginalisation)**

$$P(C_Y) = \sum_{c_Z \in \Omega_Z} P(C_Y \land c_Z).$$

(3.10)
**Proof.** For each \( c_Y \in \Omega_Y \) we have

\[
c_Y = \bigvee_{c_Z \in \Omega_Z} c_Y \land c_Z. \tag{3.11}
\]

As the disjuncts in this expression are mutually incompatible, i.e.

\[
c_Y \land c_Z \land (c_Y \land c_Z') \equiv \bot \tag{3.12}
\]

for all \( c_Z, c_Z' \in \Omega_Z, \ c_Z \neq c_Z' \), the sum of their respective probabilities equals the marginal probability of \( c_Y \).

**Proposition 3.11 (Conditioning)**

\[
P(C_Y) = \sum_{c_Z \in \Omega_Z} P(C_Y | c_Z) \cdot P(c_Z). \tag{3.13}
\]

**Proof.** Directly from the definition of conditional probability and the marginalisation property.

**Theorem 3.12 (Bayes’ theorem)**

\[
P(C_Y | C_Z) = \frac{P(C_Z | C_Y) \cdot P(C_Y)}{P(C_Z)}. \tag{3.14}
\]

**Proof.** Directly from the definition of conditional probability.

We proceed by giving definitions of *expectation* and *variance*, which play a crucial role in utility theory and decision theory. The *expected* or *mean value* of a function \( f \) under distribution \( P \) is the weighted sum of its possible values where the weights are probabilities.

**Definition 3.13 (Expected value)** Let \( f : \Omega_X \to \mathbb{R} \) be a function over the possible configurations of \( X \). The **expected value** of \( f \) under probability distribution \( P \) is defined as

\[
E_P(f) = \sum_{c_X \in \Omega_X} P(c_X) \cdot f(c_X). \tag{3.15}
\]

The variance of \( f \) is a measure of the spread around its mean; it is defined as the expected squared deviation of the values \( f(c_X), \ c_X \in \Omega_X \), from \( E_P(f) \).
Definition 3.14 (Variance) Let \( f : \Omega_X \rightarrow \mathbb{R} \) be a function over the possible configurations of \( X \). The variance of \( f \) under probability distribution \( P \) is defined as

\[
\text{var}_P(f) = E_P(g),
\]

where

\[
g(c_X) = (f(c_X) - E_P(f))^2.
\]

Note that when \( P \) is degenerate probability distribution with \( P(c_X) = 1 \) for some \( c_X \in \Omega_X \), we have that \( E_P(f) = f(c_X) \) and \( \text{var}_P(f) = 0 \).

We conclude this section by discussing the notion of entropy. Entropy was developed by Shannon and Weaver (1949) to characterise the uncertainty in a given probability distribution, which is regarded as inversely proportional to the information conveyed by that distribution. Uniform distributions are seen as conveying the least, and degenerate distributions as conveying the most information possible on the actual state of the variables involved.

Definition 3.15 (Entropy) The entropy \( H_P(Y) \) of the set \( Y \subseteq X \) in probability distribution \( P \) on \( X \) is defined as

\[
H_P(Y) = - \sum_{c_Y \in \Omega_Y} P(c_Y) \log P(c_Y).
\]

Note that \( H_P(Y) \geq 0 \) for all distributions \( P \) and sets \( Y \). The value of \( H_P(Y) \) is high when there is much uncertainty regarding the set \( Y \), i.e. when all possible configurations of \( Y \) have roughly the same marginal probability; the value of \( H_P(Y) \) is low when the marginal distribution on \( Y \) is more pronounced. If there is no uncertainty, i.e. when \( P(c_Y) = 1 \) for some \( c_Y \in \Omega_Y \), then \( H_P(Y) = 0 \).

### 3.2 Utility theory

Utility theory was formulated by Von Neumann and Morgenstern (1944) as an adjunct to their theory of games. Others soon recognised it in its own right as an important mathematical foundation for decision making under uncertainty. The central results of utility theory are that, given a number of assumptions on rational choice, a preference order on decision-making outcomes can be expressed as a real-valued function (called a utility function), and a preferred decision alternative is one that maximises the expectation of this function (Cherno and Moses, 1959; Savage, 1972; Raiffa and Schlaifer, 1961); this is called the Maximum Expected Utility criterion, or MEU criterion.

This section is divided into three subsections. Subsection 3.2.1 reviews the fundamentals of utility theory, and presents the MEU criterion. In Subsections 3.2.2 and
3.2.3, we study further characteristics of utility functions, based on attitudes towards risk and the existence of multiple objectives. We conclude in Subsection 3.2.4 with a discussion of quasi-utility functions, functions that violate the assumptions of utility theory, but may nevertheless prove useful in some circumstances.

3.2.1 The MEU criterion

Formally, in utility theory a decision is viewed as a choice from among one or more lotteries. A lottery specifies a probability distribution over “prizes” (i.e. potential outcomes of the decision) by listing them along with their respective probabilities. The decision maker receives exactly one prize, drawn using the probability distribution specified by the lottery he has chosen.

**Definition 3.16 (Lottery)** Let \( W \) be a set of variables, jointly describing the possible outcomes of a decision. The set \( \mathcal{L}(W) \) of lotteries over \( W \) is defined as the smallest set satisfying:

1. \( \Omega_W \subseteq \mathcal{L}(W) \), and
2. if \( l_1, \ldots, l_n \in \mathcal{L}(W) \), \( n \in \mathbb{N} \), \( n \geq 1 \), then \( (p_1, l_1; \ldots; p_n, l_n) \in \mathcal{L}(W) \), where \( 0 \leq p_i \leq 1 \), \( i = 1, \ldots, n \), and \( \sum_{i=1}^{n} p_i = 1 \).

A configuration \( c_W \in \Omega_W \) is also called an atomic lottery of \( \mathcal{L}(W) \). A lottery \( l = (p_1, l_1; \ldots; p_n, l_n) \), is called a simple lottery if each \( l_i \) is atomic; otherwise \( l \) is called a compound lottery. For a non-atomic lottery \( l = (p_1, l_1; \ldots; p_n, l_n) \), a pair \( (p_i, l_i) \), \( i = 1, \ldots, n \), is called a branch of \( l \) with sublottery \( l_i \).

We note that the prizes in a simple lottery need not be distinct, and neither do the prizes in sublotteries of a compound lottery. Furthermore, there are no formal restrictions on the number of branches, or on the recursion depth in a lottery, except that they are both finite. Without loss of generality, we do take non-atomic lotteries to be non-degenerate, i.e. to have some branches with a probability \( 0 < p < 1 \), and therefore involve a true gamble.

**Example 3.17** Lotteries often are depicted graphically using rooted trees. Figure 3.1a shows the simple lottery \((3/4, \text{closure} = \text{true}; 1/4, \text{closure} = \text{false})\). Figure 3.1b shows the compound lottery \((1/2, (9/10, \text{pmart} = \text{false} \land \text{death} = \text{false}; 1/10, \text{pmart} = \text{false} \land \text{death} = \text{true}); 1/2, \text{pmart} = \text{true} \land \text{death} = \text{false})\).

A compound lottery \((p_1, l_1; \ldots; p_n, l_n)\) is taken to yield the sublottery \( l_i \) with probability \( p_i \), \( i = 1, \ldots, n \), where \( l_i \) in turn is interpreted as yielding some outcome or lottery. It is easily seen that for each lottery \( l \in \mathcal{L}(W) \), there exists a unique corresponding probability distribution \( P_l \) over \( W \) that satisfies the following conditions:
3.2 Utility theory

Figure 3.1: Two lotteries.

1. if \( l \in \Omega_W \), then \( P_l(l) = 1 \) and \( P_l(c_W) = 0 \) for all \( c_W \in \Omega_W \setminus \{l\} \), and
2. if \( l = (p_1, l_1; \ldots; p_n, l_n) \), then \( P_l(c_W) = \sum_{i=1}^n p_i \cdot P_{l_i}(c_W) \) for each \( c_W \in \Omega_W \).

From a formal point of view, the set \( \mathcal{L}(W) \) is therefore equivalent to the set of possible probability distributions over \( W \). The association with gambling-style lotteries, where the outcomes typically involve winning or losing sums of money, is purely metaphoric. It does help emphasise though the assumption that the decision maker is informed about the probabilities associated with branches in a lottery and about the worths of potential outcomes.

One of the results of the correspondence between lotteries and probability distributions is that compound lotteries can be reduced to simple ones using the chain rule of probability theory (Proposition 3.9), without altering their interpretation.

Example 3.18 The compound lottery of figure 3.1b can be reduced to the simple lottery \( (9/20, \text{pmart} = \text{false} \land \text{death} = \text{false}; 1/20, \text{pmart} = \text{false} \land \text{death} = \text{true}; 1/2, \text{pmart} = \text{true} \land \text{death} = \text{false}) \) specifying the same probability distribution.

Utility theory assumes that the lotteries over a set \( W \) can be compared to each other by the decision maker as to their desirability, meaning that there exists a (subjective) preference ordering on the set \( \mathcal{L}(W) \) of all lotteries over \( W \). For two lotteries \( l, l' \in \mathcal{L}(W) \), we will write \( l \prec l' \) to denote preference of \( l' \) over \( l \) by the decision maker, \( l \succ l' \) for preference of \( l \) over \( l' \), and \( l \simeq l' \) for indifference of the decision maker between \( l \) and \( l' \). In the latter case, we will also say that \( l \) and \( l' \) are equally preferred. The symbol \( \preceq \) will be used to denote preference over or indifference between lotteries.
Definition 3.19 (Preference ordering) Let \( W \) be a set of variables and let \( \mathcal{L}(W) \) be the set of lotteries over \( W \). Then, a preference ordering on \( \mathcal{L}(W) \) is a relation \( \preceq \subseteq \mathcal{L}(W) \times \mathcal{L}(W) \) that for all \( l, l', l'' \in \mathcal{L}(W) \) satisfies the following properties:

(P1) \( \preceq \) is a linear ordering on \( \mathcal{L}(W) \); (Orderability)

(P2) if \( l \preceq l' \preceq l'' \), then there exists a \( p \) with \( 0 \leq p \leq 1 \) such that \( l' \simeq (p, l; 1-p, l'') \); (Continuity)

(P3) for all \( p \) with \( 0 \leq p \leq 1 \) we have \( l \simeq l' \) if and only if \( (p, l; 1-p, l'') \simeq (p, l'; 1-p, l'') \); (Substitutability)

(P4) if \( l \preceq l' \), then for all \( p, q \) with \( 0 \leq p, q \leq 1 \), we have \( (p, l'; 1-p, l) \preceq (q, l'; 1-q, l) \) if and only if \( p \leq q \); (Monotonicity)

(P5) for all \( p, q \) with \( 0 \leq p, q \leq 1 \), we have \( (q, (p, l; 1-p, l'); 1-q, l'') \simeq (qp, l; q(1-p), l'; 1-q, l'') \). (Decomposability)

The properties P1 through P5 are generally called the axioms of preference. The continuity axiom (P2) states that for any three lotteries \( l, l', l'' \) with \( l \preceq l' \preceq l'' \), there exists a lottery composed of \( l \) and \( l'' \) that is equally preferred to \( l' \). If \( l' \) is atomic, i.e. \( l' \in \Omega_W \), then we say that \( l' \) is the certainty equivalent of this compound lottery. Axiom P3 asserts that lotteries that are equivalent when considered alone remain equivalent as part of a larger context, and vice versa. The fourth axiom (P4, monotonicity) asserts that a decision maker prefers the lottery that offers the greater chance of receiving the better outcome. Finally, axiom P5 (decomposability) reformulates the chain rule of probability theory in lottery terms, which accords the interpretation of lotteries as probability distributions. This axiom is sometimes called “no fun in gambling,” since it prohibits one to place value on the number of steps needed to achieve an outcome.

Given a set of lotteries \( L \subseteq \mathcal{L}(W) \) and a preference ordering \( \preceq \) on \( \mathcal{L}(W) \), we say that \( l^* \in L \) is a most preferred lottery of \( L \) with respect to \( \preceq \), when \( l' \preceq l^* \) for all \( l' \in L \). Note that most preferred lotteries need not be unique, but if there are multiple most preferred lotteries in a given set, then these lotteries are all equally preferred.

For a given set of lotteries \( \mathcal{L}(W) \) and an associated preference ordering \( \preceq \), it can be shown that the axioms of preference guarantee the existence of a real-valued function over \( \mathcal{L}(W) \) that respects the decision maker’s preferences with respect to these lotteries; such a function is called a utility function.

Theorem 3.20 (Utility function) Let \( W \) be a set of variables, and let \( \preceq \) be a preference ordering on \( \mathcal{L}(W) \). Then, there exists a function \( u_{\preceq} : \mathcal{L}(W) \rightarrow \mathbb{R} \) such that for all \( l, l' \in \mathcal{L}(W) \) we have

(U1) \( u_{\preceq}(l) \leq u_{\preceq}(l') \) if and only if \( l \preceq l' \), and
(U2) if \( l = (p_1, l_1; \ldots; p_n, l_n) \), then \( u_\preceq(l) = \sum_{i=1}^n p_i \cdot u_\preceq(l_i) \).

We say that \( u_\preceq \) is a utility function for preference ordering \( \preceq \). The function value \( u_\preceq(l) \) is called the utility of lottery \( l \in \mathcal{L}(W) \) under function \( u_\preceq \).

**Proof.** See (Debreu, 1954). \( \square \)

The rules U1 and U2 are often referred to as the axioms of utility, although they describe derived, not assumed, properties of utility functions. The first axiom states that the utility of a lottery \( l' \) is greater than the utility of a lottery \( l \) if and only if \( l' \) is preferred to \( l \). The second axiom says that we can compute the utility of a compound lottery from the utilities of its constituents, using the rule of expectation from probability theory: if \( P_l \) is the probability distribution over \( W \) that corresponds to lottery \( l \), then

\[
U_\preceq(l) = E_{P_l}(u_\preceq^m),
\]

where \( u_\preceq^m \) is the function \( u_\preceq \) restricted to \( \Omega_W \), i.e. \( u_\preceq^m = u_\preceq|\Omega_W \). We refer to \( u_\preceq^m(c_W) \) as the marginal utility of outcome \( c_W \).

The axioms of preference (P1–P5) guarantee not only the existence but also the uniqueness, up to positive linear transformations, of a utility function \( u \) for a given preference ordering, (Debreu, 1954). That is, if \( u_\preceq \) is a utility function for preference ordering \( \preceq \), then so is each function \( u_\preceq' \) for which

\[
u_\preceq'(l) = a \cdot u_\preceq(l) + b,
\]

for all \( l \in \mathcal{L}(W) \), where \( a, b \in \mathbb{R} \), \( a > 0 \). The significance of this result is that the calibration commodity employed in utility functions is not the unit of outcomes’ worths, but uncertainty itself. That is, we can choose an arbitrary non-empty interval \([u_{\min}, u_{\max}] \subset \mathbb{R}\) as the range of possible utilities, and calibrate the marginal utility of outcome \( c_W \in \Omega_W \) using

\[
u_\preceq^m(c_W) = p \cdot u_{\min} + (1-p) \cdot u_{\max}
\]

if

\[
c_W \simeq (p, c_\preceq^{-}; 1-p, c_\preceq^{+}),
\]

where \( c_\preceq^{-} \) and \( c_\preceq^{+} \) are the least and most preferred outcomes, respectively. Note that the continuity axiom (P3) guarantees the existence of the compound lottery in Equation 3.22, because \( c_\preceq^{-} \preceq c_W \preceq c_\preceq^{+} \). It also follows that \( u_\preceq^m(c_\preceq^{-}) = u_{\min} \) and \( u_\preceq^m(c_\preceq^{+}) = u_{\max} \). A utility function is said to be normalised when \( u_{\min} = 0 \) and \( u_{\max} = 1 \); such a function is unique given the corresponding preference ordering \( \preceq \).

The maximum expected utility (MEU) criterion states that a decision maker, when confronted with a choice between multiple lotteries, should always choose the lottery with maximum (expected) utility.

**3.2 Utility theory**

59
Lemma 3.21 (MEU criterion) Let $W$ be a set of variables, let $\preceq$ be a preference ordering over $\mathcal{L}(W)$, and let $L \subseteq \mathcal{L}(W)$ be a non-empty set of lotteries over $W$. Then, $l^*$ is a most preferred lottery of $L$ with respect to $\preceq$ if and only if

$$l^* = \text{argmax}_l \{ u_\preceq(l) \mid l \in L \},$$

(3.23)

where $u_\preceq$ is a utility function for $\preceq$.

The MEU criterion follows from the axioms of preference and utility (see Heckerman, 1991, for an example of the proof). It is the main result of utility theory, as it provides a recipe for rational behaviour in situations of choice under uncertainty. It should be noted that such behaviour is guaranteed to yield preferred decisions (in the sense that they are consistent with one’s preferences), but not necessarily preferred outcomes: the result of a preferred decision may still turn out to be undesirable due to the uncertainty involved.

Application of utility theory to concrete decision problems under uncertainty now boils down to three steps: (1) identification of a set of variables $W$ that jointly cover all possible outcomes of decision making, (2) assessment of the marginal utility $u_\preceq^m(c_W)$ for each outcome $c_W \in \Omega_W$, and (3) choosing the decision alternative that maximises $E_P(u_\preceq^m)$, where each decision alternative is taken to explicitly or implicitly yield a probability distribution $P$ on $W$.

To facilitate the assessment of the marginal utility function $u_\preceq^m$, several qualitative characteristics of utility functions have been identified in the literature. Below, we will discuss two types of characteristics. The first type of characteristic is based on identifying a more or less objective worth, or value, with each lottery in $\mathcal{L}(W)$, and systematically comparing utilities with these values; this allows for assessing the risk attitude of the decision maker. The second type of characteristic is based on decomposition of the marginal utility function using subsets of variables from $W$; the individual variables are then referred to as attributes, and the composite function as a multiattribute utility function.

### 3.2.2 Risk attitudes

Assume that with each element $c_W$ in the outcome space $\Omega_W$ is associated a bounded numerical value $v(c_W) \in \mathbb{R}$. In a medical context, the value $v(c_W)$ often represents life-expectancy of the patient under the circumstance described by $c_W$; in an economical setting, it represents the monetary gain or loss associated with the outcome $c_W$. The expected value of lottery $l \in \mathcal{L}(W)$ is now defined as $\tilde{v}(l) = E_P(v)$, and its risk as $r(l) = \text{var}_P(v)$, where $P$ is the probability distribution on $W$ that corresponds to $l$. Note that $\tilde{v}(c_W) = v(c_W)$ and $r(c_W) = 0$ for all outcomes $c_W \in \Omega_W$, as the probability distributions associated with outcomes are degenerate. A decision maker is now said to be risk-neutral if he is indifferent between all lotteries with the same expected value. He is risk-averse if he always prefers the lottery with the smallest risk in such
3.2 Utility theory

Figure 3.2: Different risk attitudes for monotonically increasing utility. The $x$-axis represents value, and the $y$-axis represents (normalised) utility; the dots correspond to elements of $\Omega_W$, the set of all outcomes (atomic lotteries).

cases; he is risk-prone if he always prefers the largest risk. This is formalised in the following definition.

**Definition 3.22 (Risk attitude)** Let $W$ be a set of variables, let $\preceq$ be a preference ordering on $L(W)$, and let $v : \Omega_W \to \mathbb{R}$ be a value function for $W$. The decision maker is

- risk-neutral if $l' \preceq l$,
- risk-averse if $l \succ l'$, and
- risk-prone if $l \succ l'$,

for all lotteries $l, l' \in L(W)$ having $\tilde{v}(l) = \tilde{v}(l')$ and $r(l) < r(l')$. Risk aversion and risk proneness are more generally referred to as risk sensitivity.

Risk neutrality is generally viewed as optimal in situations where the objective is to maximise the cumulative result of a large sequence of similar decisions (e.g. gambling), whereas risk sensitivity is also applicable in single-decision situations (e.g. medical treatment).

The decision maker’s attitude towards risk can also be recognised from a plot of the utilities and values in the $(x, y)$-plane; see Figure 3.2 for an example. We assume here monotonically increasing utility, i.e. $u^m_{\preceq}(c_W) \geq u^m_{\preceq}(c'_W)$ if $v(c_W) \geq v(c'_W)$. This is typically the case when $v$ represents life-expectancy or monetary gain. Given a certain outcome $c_W \in \Omega_W$, a risk-neutral decision maker is indifferent between $c_W$ and any non-atomic lottery $l \in L(W)$ having $\tilde{v}(l) = v(c_W)$. That is, $u^m_{\preceq}(c_W) = u_{\preceq}(l)$; we can in fact compute all marginal utilities as geometric means once the least and most
preferred outcomes have been calibrated. Under monotonically increasing utility, risk
neutrality therefore implies that all outcomes \( c_W \in \Omega_W \) are on the line
\[
y = \frac{x - v^\text{min}}{v^\text{max} - v^\text{min}}
\]
(3.24)
where \( v^\text{min} = \min\{v(c_W) \mid c_W \in \Omega_W\} \) and \( v^\text{max} = \max\{v(c_W) \mid c_W \in \Omega_W\} \). A
risk-averse decision maker will however prefer the certain outcome \( c_W \) to lottery \( l \), as
\( r(c_W) = 0 \) and \( r(l) > 0 \). The marginal utility \( u^m_W(c_W) \) associated with \( c_W \) will there-
fore be higher than one would expect by computing the geometric mean utility \( u^g_\prec(l) \)
of lottery \( l \); all certain outcomes except the least and most preferred are therefore
located above the line of Equation 3.24 in this case. A similar argument can show
that risk proneness implies that the certain outcomes lie below the line.

Under the (generally reasonable) assumption that the decision maker is indifferent
between outcomes in \( \Omega_W \) with equal values, it is possible to define marginal utility
as a function of outcome value. That is, we define a function \( u_v : \mathbb{R} \rightarrow \mathbb{R} \) with the
restriction that
\[
u_v(E_P(v)) = E_P(u_v \circ v)
\]
(3.25)
for any joint probability distribution \( P \) on \( W \) (where ‘\( \circ \)’ denotes functional compo-
sition), and take \( u_\prec(l) = u_v(\hat{v}(l)) \) for all \( l \in \mathcal{L}(W) \). Calibration of \( u_v \) is achieved by
assessing the marginal utility of elements in outcome space directly, and by estab-
lishing the decision maker’s attitude towards risk. Note that monotonic utility now
corresponds to monotonicity of the function \( u_v \).

It can be shown that under monotonically increasing utility, a decision maker is
risk-neutral if and only if \( u_v \) is a positive linear function, he is risk-averse if and
only if his marginal utility function is concave, and he is risk-prone if and only if his
marginal utility function is convex (Keeney and Raiffa, 1976); again, see Figure 3.2 for
illustrations. Intuitively, these correspondences can be understood as follows. Given
the prospect of obtaining some value \( v, v^\text{min} < v < v^\text{max} \), a risk-neutral decision maker
is precisely twice as eager to obtain 2\( v \), a risk-averse decision maker is less than twice
as eager to obtain 2\( v \), and a risk-prone decision maker is more than twice as eager to
obtain 2\( v \). A measure of risk aversion (proneness) is found in the local risk aversion
function
\[
q(x) = -\frac{d}{dx} u_v(x).
\]
(3.26)
Concavity of \( u_v \) causes \( q \) to be positive, where higher values of \( q \) indicate stronger
risk aversion; convexity of \( u_v \) causes \( q \) to be negative, where lower values of \( q \) indi-
cate stronger risk proneness. Two utility functions are strategically equivalent (i.e.
correspond to the same preference order) if and only if they have the same local
risk aversion function (Keeney and Raiffa, 1976). Note that all risk-neutral utility
functions are strategically equivalent, as \( q(x) = 0 \) for all \( x \) for these functions.
3.2 Utility theory

3.2.3 Multiattribute utility theory

In many decision problems under uncertainty, the assessment of a preference order on the possible outcomes of decision making is hampered by the fact that there are multiple objectives involved. It is very likely these objectives conflict with each other in that the improved achievement with one objective can only be accomplished at the expense of another. For instance, in a medical setting one may wish to maximise the patient’s life-expectancy while minimising the costs of treatment. This requires, however, establishing a tradeoff between improvements in life-expectancy and the additional monetary costs.

Multiattribute utility theory (Keeney and Raiffa, 1976) is a framework for handling the value tradeoffs and uncertainties in multi-objective decision problems. In multiattribute utility theory, the outcome space \( \Omega_W \) is divided into subspaces using the variables \( w_1, \ldots, w_k \) that constitute the set \( W \); these variables are generally referred to as attributes. The basic approach is to identify regularities in the decision-making objectives that provide for decomposition of the utility function \( u^m \) to a form

\[
\text{Equation (3.27)}
\]

That is, with each attribute \( w_i, i = 1, \ldots, k \), is associated a local utility function \( u_{w_i} : \Omega_W \rightarrow \mathbb{R} \), and the utility values are subsequently combined by the function \( f \) to obtain unidimensional utility. Fundamental to the identification of regularities in the decision-maker’s objectives are the concepts of utility independence and additive independence. These concepts pertain to the influence of individual attributes or sets of attributes on the decision maker’s preferences: by restricting the possible preference orderings on \( L(W) \), they allow for simplifications in the associated utility function. Below, we assume both the global utility function \( u^m \) and all local utility functions to be normalised.

We first discuss the concept of utility independence. The underlying notion is the conditional preference ordering, which refers to the decision maker’s preferences with respect to all lotteries over a set \( Y \subseteq W \) of attributes, obtained from the preference ordering on \( L(W) \) by holding the other attributes fixed at a given value.

**Definition 3.23 (Conditional preference ordering)** Let \( W \) be a set of variables, and let \( \preceq \) be a preference ordering on \( L(W) \). Furthermore, let \( Y \subseteq W \), and let \( c_{\overline{Y}} \in \Omega_{\overline{Y}} \) be a configuration of its complementary set \( \overline{Y} = W \setminus Y \). The conditional preference ordering on \( L(Y) \) induced by \( c_{\overline{Y}} \), notation \( \preceq_{c_{\overline{Y}}} \), is defined as

\[
l_Y \preceq_{c_{\overline{Y}}} l'_Y \quad \text{if} \quad l \preceq l'
\]

for all lotteries \( l_Y, l'_Y \in L(Y) \), where \( l \) and \( l' \) are the lotteries over \( W \) obtained by replacing each \( c_Y \in \Omega_Y \) with \( c_Y \land c_{\overline{Y}} \) in \( l_Y \) and \( l'_Y \), respectively.

We now say that the set \( Y \) is utility independent of the remaining variables when the conditional preferences for lotteries over \( Y \) are the same, regardless of the configuration of \( \overline{Y} = W \setminus Y \).
Definition 3.24 (Utility independence) Let \( W \) be a set of variables, and let \( \preceq \) be a preference ordering on \( \mathcal{L}(W) \). The set \( Y \subseteq W \) is utility independent (of its complementary set \( \bar{Y} \)) when each configuration \( c_{\bar{Y}} \in \Omega_{\bar{Y}} \) induces the same conditional preference ordering \( \preceq_{\bar{Y}} \) on \( \mathcal{L}(Y) \). The variables from \( W \) are said to be mutually utility independent if every subset \( Y \) of \( W \) is utility independent of its complementary set \( \bar{Y} \).

When each configuration \( c_{\bar{Y}} \in \Omega_{\bar{Y}} \) induces the same conditional preference ordering \( \preceq_{\bar{Y}} \) on the lotteries over \( Y \), it is reasonable to speak the preference ordering on \( \mathcal{L}(Y) \) and similarly the utility function for \( Y \), independently of the other attributes. As a result, we can speak of the least and most preferred configurations \( c_{\bar{Y}} \) and \( c_{\bar{Y}}^+ \) of \( Y \), independently of the values of other attributes. When all attributes are utility independent, the following simplified form of utility function can be derived.

Theorem 3.25 (Multilinear utility) If the variables of \( W \) are mutually utility independent, then there exist functions \( u_w : \Omega_w \to \mathbb{R}, w \in W \), such that

\[
u_\preceq^m(c_W) = \sum_{Y \subseteq W, Y \neq \emptyset} k_{|Y| - 1} \prod_{w \in Y} k_w u_w(c_w) \tag{3.29}
\]

where \( k_w = u_\preceq^m(c_w^+ \wedge c_{W \setminus \{w\}}^-) \) is the weight factor for variable \( w \in W \), and \( k \) is a scaling constant that is a solution to

\[
1 + k = \prod_{w \in W} (1 + k \cdot k_w). \tag{3.30}
\]

Proof. See (Keeney and Raiffa, 1976).

The result of this theorem seems impractical as the right-hand side of Equation 3.29, which is referred to as a multilinear utility function, is rather awkward. Fortunately, there exist simplified forms of this formula that are more intuitive.

The first form of simplified multilinear utility function is obtained from the assumption of mutually additive independence of the utility attributes, (Fishburn, 1970). Under mutually additive independence, we have that

\[
u_\preceq^m(c_W^+) = \sum_{w \in W} u_\preceq^m(c_w^+ \wedge c_{W \setminus \{w\}}^-), \tag{3.31}
\]

or equally, that \( \sum_{w \in W} k_w = 1 \). It then follows that \( k = 0 \), and the multilinear utility function reduces to the following additive form:

\[
u_\preceq^m(c_W) = \sum_{w \in W} k_w u_w(c_w). \tag{3.32}
\]

Intuitively, the attributes of utility are independent additive contributors to global utility, and in optimising an individual attribute \( w \in W \), we do not have to care about
the values of other variables. The additive utility function is therefore characterised by much robustness; it is typically employed in cases where the attributes of utility represent independent factors of income.

The second form of simplified multilinear utility function is obtained from the assumption of *mutually multiplicative independence* of the utility attributes. This assumption states that for each configuration \( c_w \in \Omega_w \) of an individual attribute \( w \in W \), the ratio

\[
\frac{u_m(c_w \wedge c_{W\setminus\{w\}})}{u_m(c_w^+ \wedge c_{W\setminus\{w\}})} \tag{3.33}
\]

is the same for all configurations \( c_{W\setminus\{w\}} \) of the set \( W \setminus \{w\} \) of remaining attributes. We can then use \( u_w(c_w) \) to denote this ratio, and the multilinear utility function reduces to the following *multiplicative* form:

\[
u_m^m(C_W) = k^{|W|-1} \prod_{w \in W} k_w u_w(C_w) \tag{3.34}
\]

It models a situation where each fluctuation in local utility has a proportional effect on global utility. A multiplicative utility function is therefore not very robust: it is sometimes compared to a chain that is only as strong as its weakest link. One typically employs such a function in the case where each attribute of utility represents an independent factor of risk.

### 3.2.4 Quasi-utility functions

The axioms of preference from Definition 3.19 have not been beyond dispute. Over the years, each of the axioms P1 through P5 has been criticised, for various reasons (see Bell and Raiffa, 1988, for a discussion), and many alternatives have been formulated (e.g. Fishburn, 1988). Although the five axioms still constitute the most popular formalisation of preference under uncertainty, there are sometimes reasons to depart from them. Evaluation functions that violate one or more of the axioms of preference are called *quasi-utility functions*. In this subsection, we discuss such a function and motivate the reasons for its application.

Suppose that the decision maker indicates that he wants to minimise the uncertainty with respect to the set \( W \) of variables. That is, atomic lotteries \( c_W \in \Omega_W \) are preferred over (non-degenerate) non-atomic lotteries. More generally put, from a set of lotteries \( L \subseteq \mathcal{L}(W) \), the decision maker prefers the lottery \( l \in L \) that minimises the entropy \( H_{P_l}(W) \) of \( W \), or equivalently, maximises the negative entropy \(-H_{P_l}(W)\). An appropriate utility would therefore seem

\[
u(l) = -H_{P_l}(W). \tag{3.35}
\]

However \( u \) is not a utility function, as it violates the second axiom of utility (U2; see Theorem 3.20 on p. 58): the function regards atomic lotteries as equally (and
maximally) preferred. In a given lottery

\[ l = (p_1, c^1_W; \ldots ; p_n, c^n_W) \]  

we can therefore substitute each \( c^i_W, i = 1, \ldots , n \) with some arbitrary \( c_W \in \Omega_W \) because it is equally preferred. Then, we obtain the equivalent lottery

\[ l' = (p_1, c_W; \ldots ; p_n, c_W). \]  

However, \( l' \) yields the outcome \( c_W \) with certainty, and is therefore equivalent to the atomic lottery \( c_W \) itself. As a result, all lotteries from \( L(W) \) should be equally preferred under the preference ordering encoded by the function \( u \). This however contradicts the fact that \( u \) will yield lower (negative) values when a lottery involves much uncertainty. The function \( u \) is therefore not a utility function but a quasi-utility function.

The reasons for the invalidity of the function \( u \) as a utility function can intuitively be understood from the following principle: it is only useful to strive for more certainty when this allows for appropriate (i.e. utility-increasing) action. In a medical setting, for instance, gathering more information on the clinical state of a patient by performing a diagnostic test is considered useful only when the resulting improved diagnosis allows for better treatment. Otherwise, the adverse effects of the test such as stress, costs, or risks prove unnecessary and the test should be avoided.

Although the above principle is generally held to be sound, there may still be reasons to use a quasi-utility function that solely aims to decrease uncertainty. We distinguish three possible reasons for doing so. Firstly, information may be of prognostic value independent of its usefulness in determining treatment, (Asch et al., 1990). For instance, many patients would be willing to pay substantial sums of money simply to know more about their disease and its probable developments. Secondly, the formalisation of the decision problem may not cover all potential treatment strategies. In particular, this may be so when the formalisation aims to support only the diagnostic process, (Gorry and Barnett, 1968). Thirdly, some (complex) decision problems suffer from a combinatorial explosion in the number of possible decision policies, and can only be solved using heuristic measures. The quasi-utility function \( u \) may then be helpful in estimating the value of diagnostic tests, (Glasziou and Hilden, 1989). We will return to complexity issues in decision making at the end of the next section.

3.3 Decision-theoretic analysis

Probability theory and utility theory provide the building blocks for decision theory, the discipline that formulates the principles of rational decision making under conditions of uncertainty. The art and science of analysing real-life decision problems from these principles is called decision analysis. In this section, we provide a sketch of the field by analysing a number of prototypical decision problems under uncertainty using
decision-theoretic analysis

3.3 Decision-theoretic analysis

<table>
<thead>
<tr>
<th>Action</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>echo</td>
<td>echocardiography</td>
</tr>
<tr>
<td>med</td>
<td>medical treatment</td>
</tr>
<tr>
<td>cath</td>
<td>cardiac catheterisation</td>
</tr>
<tr>
<td>biop</td>
<td>open lung biopsy</td>
</tr>
<tr>
<td>surg</td>
<td>perform surgery</td>
</tr>
</tbody>
</table>

Table 3.2: Example decision alternatives for the VSD domain.

decision-theoretic principles. Decision-theoretic reasoning is characterised by the fact that each situation of choice is ultimately reduced to a utility-theoretic tradeoff; from a conceptual point of view, however, these choice situations may be very different. We will see that a number of concepts, pertaining to the role of actions, information and time, are significant in many decision problems. While in utility theory, these concepts are left implicit, they are typically made explicit in decision-theoretic analyses. This section aims to introduce the reader to these concepts and their role decision-theoretic reasoning.

Below, we analyse three simple decision problems: (i) making a single choice under uncertainty without prior information, (ii) making such a choice with problem-specific information, and (iii) a two-stage decision problem comprising a choice whether or not to gather information and a choice that can exploit that information. In the analyses, we confine ourselves to descriptions from sets of variables with a finite set of possible values. The choices faced by a decision maker are represented by decision variables; the values that a decision variable may take represent a mutually exclusive and exhaustive set of actions from which a choice is made at some point in the decision problem. Such actions may include the options of waiting for something to happen, inspecting or activating a given measuring device, and gathering or investing physical resources to enable future actions. We will refer to configurations of decision variables as decision alternatives. The uncertain events that are relevant to the decision problem but are beyond the (direct) control of the decision maker are represented by random variables. In the terminology of Subsection 1.2.2, these random variables describe a system under control by the decision maker, and their possible configurations represent potential system states; we will henceforth refer to them as state variables.

Example 3.26 Table 3.2 provides an example set of clinical actions that are available to a paediatric cardiologist for the management of VSD patients. Medical treatment may be used to control heart failure, and surgery may be used to close the defect. Information concerning the clinical state of the patient may be obtained by making echocardiographic images, cardiac catheterisation, and open lung biopsies. Throughout the remainder of this chapter, we will use examples with decision variables that range over this set of actions, and state variables that are taken from Table 3.1.
To visualise the various types of decision problem, we will use decision trees (Raiffa, 1968), the traditional tool for decision-theoretic analysis. A decision tree is a rooted tree that provides an explicit, graphical enumeration of the potential scenarios involved in a given decision problem; each path from root to leaf describes such a scenario. The internal nodes of the tree represent either decisions or uncertain events, and the leaf nodes represent outcomes of the decision process. As such, a decision tree highlights the structural components of the problem, i.e. (i) the alternative actions that are available to the decision maker, (ii) the events that follow from and affect these actions, and (iii) the outcomes that are associated with each possible scenario of actions and events. We note that the structure of a decision-theoretic analysis and the corresponding decision tree describe a decision problem from the perspective of the person solving the decision problem. An analysis is limited to the details of the problem relevant for choosing the optimal decision alternative(s), and ignores all other characteristics of the problem.

**Decision making in ignorance**

The simplest type of decision problem under uncertainty is where there is a single moment of choice, and there is no problem-specific information available prior to that decision: the decision maker is completely ignorant of the current state of affairs. The decision maker’s choice influences the system, and has the objective of reaching a satisfying system state, but the decision alternatives themselves are also subject to a tradeoff as they involve particular costs or risks. Using the results of Section 3.2, we will assume the decision maker’s preferences are expressed as marginal utilities for all combinations of states and choices; the underlying preference order is left implicit from now on.

**Example 3.27** Faced with a patient that has a VSD, the primary decision that needs to be made by the clinician is whether or not to submit this patient to surgery, i.e. whether or not to select the decision alternative surg. On the one hand, operating the patient will generally result in successful closure of the defect (size = null), improving the patient’s condition (shunt = none, hfail = absent), and eliminating the risks of complications such as Eisenmenger’s complex (represented by high values of the variable resis). On the other hand, a number of patients do not survive open-heart surgery (i.e. death = true).

Formally, let $d$ be a decision variable, and let $X$ be a set of state variables whose joint probability distribution depends on $d$. Let $P_{c_d}$ be the distribution on $X$ associated with decision alternative $c_d \in \Omega_d$. For convenience, we will use $p(c_X \mid c_d)$ as a shorthand notation for $P_{c_d}(c_X)$.\footnote{Note that, notwithstanding the notation, $p(c_X \mid c_d)$ is not a conditional probability as there are no marginal probabilities defined for variable $d$.} Furthermore, a marginal utility function $u$ is defined...
over $\Omega_X \times \Omega_d$. The expected utility of alternative $c_d$, denoted by $\tilde{u}(c_d)$, now equals

$$\tilde{u}(c_d) = \sum_{c_X \in \Omega_X} u(c_X, c_d) \cdot p(c_X \mid c_d).$$ (3.38)

Following the MEU criterion, a preferred, or optimal alternative $c^*_d$ is one that maximises expected utility, i.e.,

$$c^*_d = \arg\max_{c_d} \{ \tilde{u}(c_d) \mid c_d \in \Omega_d \}. \quad (3.39)$$

Note that there need not be a unique preferred alternative. The term regret refers to the expected utility that is `missed' by choosing a sub-optimal decision alternative.

**Definition 3.28 (Regret)** The regret of decision $c_d$ is defined as $\tilde{u}(c_d) - \tilde{u}(c^*_d)$, where $c^*_d$ is an optimal alternative.

It is easily seen that regret is non-positive, and the lower the regret, the worse the decision.

The decision problem described above is represented by the decision tree of Figure 3.3. Decisions are depicted by squares, called decision nodes, and labelled with the name of the corresponding decision variable; the branches emanating from a decision node correspond to the possible decision alternatives (in this case $c^1_d, \ldots, c^n_d$). Uncertain events are depicted by circles, called chance nodes, and labelled with the name of
the corresponding set of random variables. In general, there may be multiple nodes
in a decision tree representing the same decision or uncertain event (albeit under
different circumstances); in this case, there are \( n \) chance nodes, each representing
the uncertain event with possible outcomes in the universe \( \Omega_X \) of the set \( X \).
The branches emanating from a chance node correspond to these possible outcomes (in
this case \( c_X^1, \ldots, c_X^n \)). With each of the branches is associated the probability
that the event takes place in that context; for instance, the upper branch emanating from
the middle chance node has associated probability \( p(c_X^1 | c_d) \). These probabilities are
not shown in the figure. Leaf nodes of the tree represent potential outcomes of the
decision-making process; each leaf node is labelled with the utility that corresponds
to the sequence of decisions and events along the unique path to that node starting
at the root.

As is seen from the figure, the decision-tree representation of decision problems re-
jects the utility-theoretic conception of decisions as choices between lotteries. By
\textit{evaluating or solving} a decision tree is meant finding an optimal \textit{policy} for the tree,
i.e. selecting an optimal branch for each decision node in the tree. In this case, there
is only one decision node, and an optimal branch is one which corresponds to an
alternative that satisfies the condition in Equation 3.39.

We conclude the analysis of this decision problem by noting that it is well possible
that one or more state variables are unaffected by the decision, i.e. that there exists
a subset \( Y \subseteq X \) for which

\[
P_{c_d}(C_Y) = P_{c_d}(C_Y) \tag{3.40}
\]

for all \( c_d, c_d' \in \Omega_d \). We can then regard these variables as receiving their values
prior to decision making. Recall, however, that a decision analysis solely reflects
the perspective of the decision maker. The circumstance described therefore has no
consequences for the analysis of the decision problem, as we have assumed the decision
maker to be ignorant of the values of state variables when making the decision.
Below, we will lift the assumption of complete ignorance, and analyse the effect of
observations.

\textbf{Decision making with prior observations}

It often comes about that a decision maker has some information regarding the state
of the system to which his decisions pertain. In our formalisation, this information
consists of configurations of one or more state variables; we will refer to these configu-
rations as \textit{evidence}. We will first discuss \textit{case parameters}, the special type of evidence
that is received prior to making the (first) decision, and turn later to observations
that are received as a result of specific choices by the decision maker. In a medi-
cal setting, case parameters typically consist of patient-specific information such as
personal and historical data, symptoms, and findings from physical examination.

\textbf{Example 3.29} In the VSD example, case parameters consist of externally visible
signs and symptoms, typically caused by heart failure and shunting: shortness of breath, feeding and growing problems (comprised in the variable symp in Table 3.1), and occurrence of central cyanosis (cyan = true).

Formally, let as before $X$ be a set of random variables, let $d$ be a decision variable, and let $u$ be a marginal utility function over $\Omega_X \times \Omega_d$. Now, assume that $Y \subseteq X$ is a set of state variables that receive their values prior to decision making, and are observed at that time. We use $p_Y(c_Y)$ to denote the marginal probability of configuration $c_Y$, i.e. $p_Y(c_Y) = P_{c_d}(c_Y)$ for any $c_d \in \Omega_d$, and use
\[
p_{X\setminus Y}(c_{X\setminus Y} | c_Y, c_d) = P_{c_d}(c_{X\setminus Y} | c_Y)
\]
(3.41) to denote the conditional probability of configuration $c_{X\setminus Y}$ given $c_Y$ when choosing alternative $c_d$; the derived conditional distribution $p_{X\setminus Y}$ can be seen as describing the effects of the variables from $Y$ on the variables from $X \setminus Y$ under the various decision alternatives.

**Example 3.30** Let $X = \{\text{shunt}, \text{hfail}\}$ and consider the decision to administer cardiac glycosides to enhance the strength of myocardial contraction. This decision will not affect the shunt, but it does in general reduce heart failure. Taking $Y = \{\text{shunt}\}$, the conditional probability $p_{X\setminus Y}$ describes the effects of shunting on heart failure under both decision alternatives. In general, larger shunt sizes will increase the risk of (severe) heart failure, but the risk will decrease under treatment with glycosides.

The decision tree depicted in Figure 3.4 models the decision problem with prior observations. The root of the tree is a chance node representing the observed set $Y$ of state variables, and is followed by decision $d$. The configuration of $Y$, which is then known to the decision maker, is used to optimise the decision. Generally speaking, all variables preceding a decision node in a decision tree are assumed to be known when making the decision; this may also include earlier decisions and observations that result from specific choices. Chance nodes not followed by a decision node, in this case representing the set $X \setminus Y$, either represent attributes of the system that remain hidden from observation, or model an uncertain event in the future. Recall that with each branch emanating from a chance node is associated the conditional probability of the uncertain event given the history of past decisions and events along the path that leads from the root to the node; here, the history consists of configurations of the set $Y \cup \{d\}$.

The form of the tree stresses the fact that with each observation from $\Omega_Y$, we face a different subproblem, and each of these subproblems may have its own optimal solution. To formulate solutions to observation-dependent subproblems, we use decision functions.

**Definition 3.31 (Decision function)** Let $W$ be a set of variables, and let $d$ be a decision variable. A decision function for $d$ is a function $\delta : \Omega_W \rightarrow \Omega_d$. Configu-
rations of $W$ are called inputs to the decision function $\delta$. The set of all decision functions for $d$ with inputs from $\Omega_W$ is written $\Delta_{W-d}$.

Both configurations of decision variables and random variables can serve as inputs to a decision function. Note that when $W = \emptyset$, then the function simply picks an alternative from $\Omega_d$; such functions can therefore be used for decisions without prior observations.

The problem faced by the decision maker is to solve the decision tree by selecting an optimal branch for each of the tree's decision nodes, or equivalently, to choose the optimal decision function for $d$ with inputs from $\Omega_Y$. We define the conditional expected utility of decision $c_d \in \Omega_d$ given observation $c_Y$, notation $\tilde{u}(c_d \mid c_Y)$, as

$$
\tilde{u}(c_d \mid c_Y) = \sum_{c_{X \setminus Y} \in \Omega_{X \setminus Y}} u(c_Y \land c_{X \setminus Y}, c_d) \cdot p_{X \setminus Y}(c_{X \setminus Y} \mid c_Y, c_d),
$$

(3.42)
and the (unconditional) expected utility $\hat{u}(\delta)$ of decision function $\delta$ as

$$
\hat{u}(\delta) = \sum_{c_Y \in \Omega_Y} \hat{u}(\delta(c_Y) \mid c_Y) \cdot p_Y(c_Y).
$$

The optimal decision function is

$$
\delta^* = \arg\max_{\delta \in \Delta_{Y-d}} \hat{u}(\delta),
$$

or equivalently, the function for which

$$
\delta^*(c_Y) = \arg\max_{c_d \in \Omega_d} \{\hat{u}(c_d \mid c_Y) \mid c_d \in \Omega_d\}.
$$

for all $c_Y \in \Omega_Y$. The latter formulation reflects the fact that there are $|\Omega_Y|$ different subproblems, for each of which an optimal solution has to be found. We remark that the notion of regret from Definition 3.28 can be defined analogously for decision functions.

The following notion is due to Howard (1966).

**Definition 3.32 (Value of information)** The expected value of information of the set $Y$, notation $\text{EVI}(Y)$, is defined as

$$
\text{EVI}(Y) = \hat{u}(\delta^*) - \hat{u}(c_d^*),
$$

where $\hat{u}(\delta^*)$ is the optimal decision function with inputs from $\Omega_Y$, and $c_d^*$ is the optimal decision alternative without prior information.

If $Y$ is the largest set of state variables uninfluenced by the decision, this means that an observed configuration of $Y$ represents all there is to know of the problem prior to decision making. We then refer to EVI$(Y)$ as the expected value of perfect information (EVPI).

Proposition 3.33 now expresses the decision-theoretic principle that a decision maker should always use as much of the available information as possible, as long as the information is freely available.

**Proposition 3.33** The expected value of information $\text{EVI}(Y)$ of any set $Y$ is non-negative.

**Proof.** Let $c_d^*$ as before be the decision alternative that expectedly optimises utility when there is no prior information, and let $\delta^b$ be the constant decision function that selects $c_d^*$ whatever evidence on the set $Y$ is received (where $b$ refers to the ‘blindness’ of this decision function). From Equation 3.43, we have that

$$
\hat{u}(\delta^b) = \sum_{c_Y \in \Omega_Y} \hat{u}(c_d^* \mid c_Y) \cdot p_Y(c_Y) = \hat{u}(c_d^*).
$$
Now, as $\delta^*$ maximises $\tilde{u}(\delta^*(c_Y) \mid c_Y)$ for each $c_Y \in \Omega_Y$ (Equation 3.45), we have that
\[
\tilde{u}(\delta^*(c_Y) \mid c_Y) \geq \tilde{u}(\delta^b(c_Y) \mid c_Y) \tag{3.48}
\]
for all $c_Y \in \Omega_Y$, and therefore
\[
\tilde{u}(\delta^*) \geq \tilde{u}(\delta^b), \tag{3.49}
\]
so
\[
\tilde{u}(\delta^*) \geq \tilde{u}(c_d^*). \tag{3.50}
\]

It appears from the proof that the expected value of information of $Y$ can be viewed as the negative regret of ignoring the evidence by using the blind decision function $\delta^b$.

We note that it is possible to re-formulate a decision problem with prior observations as a singular choice without prior observations. Let $d'$ be a new decision variable, taking values from $\Delta_{Y \rightarrow d}$, and define, for each alternative $\delta \in \Delta_{Y \rightarrow d}$,
\[
p'_{X \setminus Y}(c_{X \setminus Y} \mid d' = \delta) = \sum_{c_Y \in \Omega_Y} p_Y(c_Y) \cdot p_{X \setminus Y}(c_{X \setminus Y} \mid c_Y \land \delta(c_X)), \tag{3.51}
\]
and
\[
\tilde{u}'(d' = \delta) = \sum_{c_{X \setminus Y} \in \Omega_{X \setminus Y}} u(c_Y \land c_{X \setminus Y}, \delta(c_Y)) \cdot p'_{X \setminus Y}(c_{X \setminus Y} \mid d' = \delta). \tag{3.52}
\]
We then have that $\tilde{u}'(d' = \delta) = \tilde{u}(\delta)$. In this formulation, the corresponding decision tree again has the form of the tree depicted in Figure 3.3. The number of branches emanating from the root decision will however be much larger, as there exist $n^z$ possible decision functions for $d$ with inputs from $Y$ if $n = |\Omega_d|$ is the number of decision alternatives and $z = |\Omega_Y|$ is the number of possible observations. This is much more than the number of $n$ decision alternatives when there is no prior evidence. We conclude that the possibility to observe prior evidence does not introduce a fundamental difference to the type of tradeoff involved, but there is a substantial increase in complexity of the problem.

**Deciding upon intermediate observations**

In the previous decision problem, we assumed that evidence was freely available, and therefore provided a guaranteed increase in expected utility. In many decision problems, evidence concerning the system state is however not freely available but comes forth as a result of specific choices made during the decision process; we will refer to such choices as *test decisions*. 
Definition 3.34 (Test decision) A test decision is a decision with the objective to gather evidence.

On the one hand, the evidence obtained from test decisions can be used to optimise subsequent decisions, but on the other hand the conduction of tests will generally involve certain costs or risks. In a decision-theoretic analysis, these costs and risks are discounted as a decrease in utility. The question is therefore whether this immediate decrease in utility is justified by the expected increase in utility that results from improved future decision making.

Example 3.35 A cardiologist may choose to make an echographic image of the heart of a VSD patient (the decision alternative echo), providing information about the size of the VSD (the state variable size). In addition, cardiac catheterisation (cath) may be used to obtain evidence on shunting and vascular resistances (variables shunt and resis), and the state of the pulmonary arterioles (pmart) can be examined by open lung biopsy (the alternative biop). Each of these types of evidence is helpful in assessing the severity of disease, and therefore in deciding upon the need for surgery.

We formalise this situation as follows. Let $d_1, d_2$ be decision variables, where $d_1$ denotes a test decision; it takes one of the values test and no_test. When $d_1 = \text{test}$, the decision maker receives evidence on the set $Y \subseteq X$ of state variables; otherwise, he has no information on the state of the system. The probability distribution $P$ on $X$ is not influenced by $d_1$ and therefore only parametrised by $d_2$; the utility function $u$ depends on $X$, $d_1$, and $d_2$.

The decision tree that corresponds to this problem is shown in Figure 3.5. As appears from the figure, the test decision is basically a choice between a decision with prior observations (represented by the upper half of the tree) and the same decision without observations (represented by the lower half of the tree). Also note that in contrast with the trees of Figures 3.3 and 3.4, the tree of Figure 3.5 is not symmetrical: the two halves of the tree are structurally different. We will refer to this phenomenon, which is induced by test decisions, as informational asymmetry. Informational asymmetry abounds in all decision problems that involve test decisions, and is therefore a significant phenomenon in decision-theoretic representation and reasoning.

Due to the asymmetric nature of the problem, the analysis is also split into two parts: one part where it is decided to perform the test, and one where it is decided to skip it. First, suppose that $d_1 = \text{test}$ is selected. This means that information on the configuration of $Y$ becomes available, and we can use this information to optimise decision $d_2$. That is, we will then choose a decision function $\delta \in \Delta_{Y,d_2}$ for $d_2$, and the expected utility of this function is

$$
\bar{u}(\delta \mid d_1 = \text{test}) = \sum_{c_Y \in \partial_Y} p_Y(c_Y) \cdot \bar{u}(\delta(c_Y) \mid c_Y, d_1 = \text{test}),
$$

(3.53)
Figure 3.5: Decision tree with a test decision.
where the expected utility \( \hat{u}(c_d \mid c_Y, d_1 = \text{test}) \) of decision alternative \( c_{d_2} \) given \( d_1 = \text{test} \) and evidence \( c_Y \) is defined as

\[
\hat{u}(c_d \mid c_Y, d_1 = \text{test}) = \sum_{c_{X|Y} \in \Omega_{X|Y}} u(c_Y \land c_{X|Y}, d_1 = \text{test}, c_{d_2}) \cdot p_{X|Y}(c_{X|Y} \mid c_Y \land c_{d_2}). \tag{3.54}
\]

Note that \( \delta(c_Y) \) in Equation 3.53 yields such a decision alternative for decision \( d \), based on the evidence \( c_Y \). The (maximum) expected utility of making the observation is therefore

\[
\hat{u}(d_1 = \text{test}) = \max \{ \hat{u}(\delta \mid d_1 = \text{test}) \mid \delta \in \Delta_{Y \rightarrow d_2} \}. \tag{3.55}
\]

In contrast, when we decide to choose \( \text{no}_\text{test} \) for decision \( d_1 \), then no information about \( Y \) comes available and all we can do is select the alternative from \( d_2 \) that is expected to be optimal given the prior distribution on \( X \). The expected utility of alternative \( c_{d_2} \in \Omega_{d_2} \) equals

\[
\hat{u}(c_{d_2} \mid d_1 = \text{no}_\text{test}) = \sum_{c_X \in \Omega_X} p_X(c_X) \cdot u(c_X, d_1 = \text{no}_\text{test}, c_{d_2}), \tag{3.56}
\]

and the (maximum) expected utility of not observing is

\[
\hat{u}(d_1 = \text{no}_\text{test}) = \max \{ \hat{u}(c_{d_2} \mid d_1 = \text{no}_\text{test}) \mid c_{d_2} \in \Omega_{d_2} \}. \tag{3.57}
\]

If \( \hat{u}(d_1 = \text{test}) > \hat{u}(d_1 = \text{no}_\text{test}) \), then we perform the test and apply the optimal decision function for variable \( d_2 \) afterwards; otherwise, we do not make the observation and select the optimal alternative from \( \Omega_{d_2} \).

The difference \( \hat{u}(d_1 = \text{test}) - \hat{u}(d_1 = \text{no}_\text{test}) \) in maximum expected utilities between testing and not testing is sometimes referred to as the expected test value. Note that when the test decision does not affect utility, i.e. when

\[
u(C_X, d_1 = \text{test}, C_{d_2}) = u(C_X, d_1 = \text{no}_\text{test}, C_{d_2}) \tag{3.58}\]

then the expected test value equals the expected value of information \( \text{EVI}(Y) \) of the set \( Y \). If follows from Proposition 3.33 that the expected test value then be non-negative, and it is therefore recommended to perform it. In most practical settings, however, information will not be available ‘for free’. That is, we usually have that

\[
u(C_X, d_1 = \text{test}, C_{d_2}) < u(C_X, d_1 = \text{no}_\text{test}, C_{d_2}), \tag{3.59}\]

and there is a tradeoff between loss in utility on the one hand and gain in information that may compensate that loss on the other hand.

We conclude this subsection with a few notes on the generalised problem where multiple tests are available that can be conducted serially. An example of this situation is found in the differential diagnosis task described in Subsection 2.2.2, where the
physician repeatedly selects diagnostic tests until sufficient information is available to choose appropriate therapy. Firstly, if the tests can be performed in any order, their expected values depend on the actual order in which the tests are performed, and can therefore not be assessed in isolation. The problem becomes highly combinatorial as in principle, any sequence of tests should be considered.

Example 3.36 Suppose that a paediatric cardiologist suspects that a VSD patient suffers from pulmonary hypertension due to pulmonary arteriopathy. If his suspicion is correct, surgical closure of the VSD will worsen the patient’s condition. He may now consider the tests cath (cardiac catheterisation) to inspect intra-cardiac flows and pressures and biop (open lung biopsy) to inspect the pulmonary arterioles directly. Both tests are invasive but not completely reliable. When a test is chosen and the results are equivocal, additional testing is required.

Secondly, the existence of multiple tests will generally induce relevantial asymmetry in the formal analysis of the problem. This can be understood as follows. Suppose that \( d_1, \ldots, d_k \) denote test decisions, where each \( d_i, i = 1, \ldots, k \), takes values from an arbitrary set of tests \( A \). Now, if \( a \in A \) denotes a test that is guaranteed to provide evidence that is free of measurement errors, then selecting \( d_i = a, i = 1, \ldots, k - 1 \), rules out the necessity to consider test \( a \) for any future decision \( d_j, j = i + 1, \ldots, k \), as conducting the test would not provide any further information: the option \( a \) has become irrelevant. Relevantial asymmetry is reminiscent of informational asymmetry, but pertains to decision variables instead of state variables; it generally occurs in problems where decisions (not necessarily pertaining to tests) may be repeated.

Example 3.37 Consider once more the above example. If we assume that an open lung biopsy provides reliable information on the state of the pulmonary arterioles, then it is unnecessary to repeat this test once it has been performed.

3.4 Discussion

In the previous sections, we have discussed the formal foundations for decision making under uncertainty within the decision-theoretic paradigm. It was described how the synthesis of probability theory and utility theory provides a framework for analysing various types of choice under uncertain conditions. Within this framework, probability theory serves to formalise the reasoning about uncertain events, whereas utility theory provides the guidelines for rational choice under uncertainty. The principal rule in any circumstance is that a person facing a decision should make the choice that is expectedly optimal with respect to his preferences.

Although decision theory is a normative theory as it prescribes the preferred behaviour of a decision maker, in part it also has a descriptive character (Kyburg, 1991).
One would expect the goal of a normative theory of decision to provide rules that would unconditionally optimise one’s preferences. This is however not the goal of decision theory because we take it for granted that no theory can do this: it is part of the human condition that we cannot predict the future perfectly, and therefore cannot choose the course of action that will in fact maximise our satisfactions. So, what we take as a normative theory, depends on what we take to be possible for human agents – a matter of descriptive rather than normative character.

The implications of this ‘realistic’ character of decision theory are more than purely philosophical. The complexity of most practical decision problems stems from the circumstance that there is a large number of decisions involved, and each potential policy for making these decisions induces a multitude of possible scenarios. As we cannot hope to fulfil our preferences completely, we are ultimately forced to evaluate all policies by inspecting each possible scenario. In other words, practical application of decision theory is hampered by the facts that decisions can often not be made in isolation but instead have to be evaluated in the context of other decisions, and that all possible consequences of decisions have to be taken into account in these evaluations.

To illustrate the highly combinatorial nature of decision problems, consider a problem that involves $k$ subsequent decisions. If $z$ is an upper bound on the number of different observations that may be obtained before each decision, and $n$ is the maximum number of alternatives for each decision, then there exist $O(n^z)$ different decision functions for the $k$th decision. In a decision-tree analysis of $k$ subsequent decisions, the tree will have $O(n^k z^k)$ nodes. Evaluation of all decision-making policies, or equivalently, solving the decision tree, is therefore hopelessly intractable when $k$ is large. It should be noted though, that the complexity of the problem does not alter the nature of the trade-off between decision alternatives on utility-theoretic grounds: there is no fundamental difference between a simple choice between decision alternatives in a single-decision problem and a choice between policies in a problem with multiple decisions.

The decision-theoretic analyses in this chapter have been illustrated with decision trees. Decision trees are frequently used in the field of clinical decision analysis (Weinstein and Fineberg, 1980; Pauker and Kassirer, 1987) as they provide an intuitive representation of decision problems and can easily be constructed in cooperation with experienced clinicians. They are however not suited as a knowledge-representation formalism for automated reasoning systems. The reasons for this are threefold. First, decision trees describe decision problems from the viewpoint of the decision maker while leaving most of the underlying knowledge of the problem domain implicit. Second, as we discussed above, the size of a decision tree grows exponentially in the size of the problem because a tree explicitly enumerates all possible decision-making policies. Third, they can only be used to solve a single problem case, whereas

---

2We assume that there are no concurrent actions. This is sometimes referred to as the single decision maker assumption.
a knowledge-based system would typically cover a range of problem cases within a given domain.

While decision trees are impractical as a knowledge-representation formalism for decision-theoretic reasoning systems, there exists several other representation formalisms that adhere to the decision-theoretic perspective, and may be used as a basis for intelligent reasoning. These formalisms are discussed in the next chapter.