CHAPTER 15

MAKING SIMPLE DECISIONS

In which we see how an agent should make decisions so that it gets what it wants in an uncertain world—at least as much as possible and on average.

In this chapter, we fill in the details of how utility theory combines with probability theory to yield a decision-theoretic agent—an agent that can make rational decisions based on what it believes and what it wants. Such an agent can make decisions in contexts in which uncertainty and conflicting goals leave a logical agent with no way to decide. A goal-based agent has a binary distinction between good (goal) and bad (non-goal) states, while a decision-theoretic agent assigns a continuous range of values to states, and thus can more easily choose a better state even when no best state is available.

Section 15.1 introduces the basic principle of decision theory: the maximization of expected utility. Section 15.2 shows that the behavior of a rational agent can be modeled by maximizing a utility function. Section 15.3 discusses the nature of utility functions in more detail, and in particular their relation to individual quantities such as money. Section 15.4 shows how to handle utility functions that depend on several quantities. In Section 15.5, we describe the implementation of decision-making systems. In particular, we introduce a formalism called a **decision network** (also known as an **influence diagram**) that extends Bayesian networks by incorporating actions and utilities. Section 15.6 shows how a decision-theoretic agent can calculate the value of acquiring new information to improve its decisions.

While Sections 15.1–15.6 assume that the agent operates with a given, known utility function, Section 15.7 relaxes this assumption. We discuss the consequences of preference uncertainty on the part of the machine—the most important of which is deference to humans.

15.1 Combining Beliefs and Desires under Uncertainty

We begin with an agent that, like all agents, has to make a decision. It has available some actions a. There may be uncertainty about the current state, so we'll assume that the agent assigns a probability $P(s)$ to each possible current state s. There may also be uncertainty about the action outcomes; the transition model is given by $P(s' \mid s, a)$, the probability that action a in state s reaches state s'. Because we're primarily interested in the outcome s', we'll also use the abbreviated notation $P(\text{RESULT}(a) = s')$, the probability of reaching s' by doing a in the current state, whatever that is. The two are related as follows:

$$P(\text{RESULT}(a) = s') = \sum_s P(s) P(s' \mid s, a) .$$

Decision theory, in its simplest form, deals with choosing among actions based on the desirability of their *immediate* outcomes; that is, the environment is assumed to be episodic in the

sense defined on page 63. (This assumption is relaxed in Chapter 16.) The agent's preferences are captured by a **utility function**, $U(s)$, which assigns a single number to express the desirability of a state. The **expected utility** of an action given the evidence, $EU(a)$, is just the average utility value of the outcomes, weighted by the probability that the outcome occurs:

Utility function
Expected utility

$$EU(a) = \sum_{s'} P(\text{RESULT}(a) = s')\, U(s')\,. \tag{15.1}$$

The principle of **maximum expected utility (MEU)** says that a rational agent should choose the action that maximizes the agent's expected utility:

$$action = \operatorname*{argmax}_a EU(a)\,.$$

In a sense, the MEU principle could be seen as a prescription for intelligent behavior. All an intelligent agent has to do is calculate the various quantities, maximize utility over its actions, and away it goes. But this does not mean that the AI problem is *solved* by the definition!

The MEU principle *formalizes* the general notion that an intelligent agent should "do the right thing," but does not *operationalize* that advice. Estimating the probability distribution $P(s)$ over possible states of the world, which folds into $P(\text{RESULT}(a) = s')$, requires perception, learning, knowledge representation, and inference. Computing $P(\text{RESULT}(a) = s')$ itself requires a causal model of the world. There may be many actions to consider, and computing the outcome utilities $U(s')$ may itself require further searching or planning because an agent may not know how good a state is until it knows where it can get to from that state. An AI system acting on behalf of a human may not know the human's true utility function, so there may be uncertainty about U. In summary, decision theory is not a panacea that solves the AI problem—but it does provide the beginnings of a basic mathematical framework that is general enough to define the AI problem.

The MEU principle has a clear relation to the idea of performance measures introduced in Chapter 2. The basic idea is simple. Consider the environments that could lead to an agent having a given percept history, and consider the different agents that we could design. *If an agent acts so as to maximize a utility function that correctly reflects the performance measure, then the agent will achieve the highest possible performance score (averaged over all the possible environments).* This is the central justification for the MEU principle itself. While the claim may seem tautological, it does in fact embody a very important transition from the external performance measure to an internal utility function. The performance measure gives a score for a history—a sequence of states. Thus it is applied retrospectively after an agent completes a sequence of actions. The utility function applies to the very next state, so it can be used to guide actions step by step.

15.2 The Basis of Utility Theory

Intuitively, the principle of Maximum Expected Utility (MEU) seems like a reasonable way to make decisions, but it is by no means obvious that it is the *only* rational way. After all, why should maximizing the *average* utility be so special? What's wrong with an agent that maximizes the weighted sum of the cubes of the possible utilities, or tries to minimize the worst possible loss? Could an agent act rationally just by expressing preferences between states, without giving them numeric values? Finally, why should a utility function with the required properties exist at all? We shall see.

15.2.1 Constraints on rational preferences

These questions can be answered by writing down some constraints on the preferences that a rational agent should have and then showing that the MEU principle can be derived from the constraints. We use the following notation to describe an agent's preferences:

$A \succ B$ the agent prefers A over B.

$A \sim B$ the agent is indifferent between A and B.

$A \succsim B$ the agent prefers A over B or is indifferent between them.

Now the obvious question is, what sorts of things are A and B? They could be states of the world, but more often than not there is uncertainty about what is really being offered. For example, an airline passenger who is offered "the pasta dish or the chicken" does not know what lurks beneath the tinfoil cover.[1] The pasta could be delicious or congealed, the chicken juicy or overcooked beyond recognition. We can think of the set of outcomes for each action as a **lottery**—think of each action as a ticket. A lottery L with possible outcomes $S_1, \ldots, S_n$ that occur with probabilities $p_1, \ldots, p_n$ is written

Lottery

$$L = [p_1, S_1;\ p_2, S_2;\ \ldots\ p_n, S_n]\,.$$

In general, each outcome S_i of a lottery can be either an atomic state or another lottery. The primary issue for utility theory is to understand how preferences between complex lotteries are related to preferences between the underlying states in those lotteries. To address this issue we list six constraints that we require any reasonable preference relation to obey:

Orderability

- **Orderability**: Given any two lotteries, a rational agent must either prefer one or else rate them as equally preferable. That is, the agent cannot avoid deciding. As noted on page 412, refusing to bet is like refusing to allow time to pass.

  $$\text{Exactly one of } (A \succ B),\ (B \succ A),\ \text{or } (A \sim B) \text{ holds.}$$

Transitivity

- **Transitivity**: Given any three lotteries, if an agent prefers A to B and prefers B to C, then the agent must prefer A to C.

  $$(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)\,.$$

Continuity

- **Continuity**: If some lottery B is between A and C in preference, then there is some probability p for which the rational agent will be indifferent between getting B for sure and the lottery that yields A with probability p and C with probability $1-p$.

  $$A \succ B \succ C \Rightarrow \exists p\ [p, A;\ 1-p, C] \sim B\,.$$

Substitutability

- **Substitutability**: If an agent is indifferent between two lotteries A and B, then the agent is indifferent between two more complex lotteries that are the same except that B is substituted for A in one of them. This holds regardless of the probabilities and the other outcome(s) in the lotteries.

  $$A \sim B \Rightarrow [p, A;\ 1-p, C] \sim [p, B; 1-p, C]\,.$$

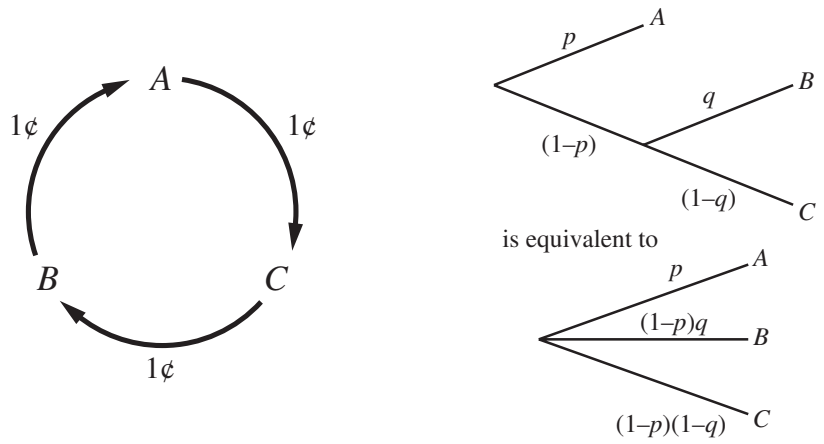
  This also holds if we substitute $\succ$ for $\sim$ in this axiom.

Monotonicity

- **Monotonicity**: Suppose two lotteries have the same two possible outcomes, A and B. If an agent prefers A to B, then the agent must prefer the lottery that has a higher probability for A (and vice versa).

  $$A \succ B \Rightarrow (p > q \Leftrightarrow [p, A;\ 1-p, B] \succ [q, A;\ 1-q, B])\,.$$

[1] We apologize to readers whose local airlines no longer offer food on long flights.



(a) (b)

Figure 15.1 (a) Nontransitive preferences $A \succ B \succ C \succ A$ can result in irrational behavior: a cycle of exchanges each costing one cent. (b) The decomposability axiom.

- **Decomposability**: Compound lotteries can be reduced to simpler ones using the laws of probability. This has been called the "no fun in gambling" rule: as Figure 15.1(b) shows, it compresses two consecutive lotteries into a single equivalent lottery.[2] Decomposability

$$[p,A;\ 1-p,[q,B;\ 1-q,C]] \sim [p,A;\ (1-p)q,B;\ (1-p)(1-q),C]\,.$$

These constraints are known as the axioms of utility theory. Each axiom can be motivated by showing that an agent that violates it will exhibit patently irrational behavior in some situations. For example, we can motivate transitivity by making an agent with nontransitive preferences give us all its money. Suppose that the agent has the nontransitive preferences $A \succ B \succ C \succ A$, where A, B, and C are goods that can be freely exchanged. If the agent currently has A, then we could offer to trade C for A plus one cent. The agent prefers C, and so would be willing to make this trade. We could then offer to trade B for C, extracting another cent, and finally trade A for B. This brings us back where we started from, except that the agent has given us three cents (Figure 15.1(a)). We can keep going around the cycle until the agent has no money at all. Clearly, the agent has acted irrationally in this case.

15.2.2 Rational preferences lead to utility

Notice that the axioms of utility theory are really axioms about preferences—they say nothing about a utility function. But in fact from the axioms of utility we can derive the following consequences (for the proof, see von Neumann and Morgenstern, 1944):

- **Existence of Utility Function**: If an agent's preferences obey the axioms of utility, then there exists a function U such that $U(A) > U(B)$ if and only if A is preferred to B, and $U(A) = U(B)$ if and only if the agent is indifferent between A and B. That is,

$$U(A) > U(B) \;\Leftrightarrow\; A \succ B \qquad \text{and} \qquad U(A) = U(B) \;\Leftrightarrow\; A \sim B\,.$$

[2] We can account for the enjoyment of gambling by encoding gambling events into the state description; for example, "Have $10 and gambled" could be preferred to "Have $10 and didn't gamble."

- **Expected Utility of a Lottery**: The utility of a lottery is the sum of the probability of each outcome times the utility of that outcome.

$$U([p_1, S_1; \ldots; p_n, S_n]) = \sum_i p_i U(S_i)\,.$$

In other words, once the probabilities and utilities of the possible outcome states are specified, the utility of a compound lottery involving those states is completely determined. Because the outcome of a nondeterministic action is a lottery, it follows that an agent can act rationally—that is, consistently with its preferences—only by choosing an action that maximizes expected utility according to Equation (15.1).

The preceding theorems establish that (assuming the constraints on rational preferences) a utility function *exists* for any rational agent. The theorems do not establish that the utility function is *unique*. It is easy to see, in fact, that an agent's behavior would not change if its utility function $U(S)$ were transformed according to

$$U'(S) = aU(S) + b\,, \tag{15.2}$$

where a and b are constants and $a > 0$; a positive affine transformation.[3] This fact was noted in Chapter 6 (page 213) for two-player games of chance; here, we see that it applies to all kinds of decision scenarios.

Value function
Ordinal utility function

As in game-playing, in a deterministic environment an agent needs only a preference ranking on states—the numbers don't matter. This is called a **value function** or **ordinal utility function**.

It is important to remember that the existence of a utility function that describes an agent's preference behavior does not necessarily mean that the agent is *explicitly* maximizing that utility function in its own deliberations. As we showed in Chapter 2, rational behavior can be generated in any number of ways. A rational agent might be implemented with a table lookup (if the number of possible states is small enough).

By observing a rational agent's behavior, an observer can learn about the utility function that represents what the agent is actually trying to achieve (even if the agent doesn't know it). We return to this point in Section 15.7.

15.3 Utility Functions

Utility functions map from lotteries to real numbers. We know they must obey the axioms of orderability, transitivity, continuity, substitutability, monotonicity, and decomposability. Is that all we can say about utility functions? Strictly speaking, that is it: an agent can have any preferences it likes. For example, an agent might prefer to have a prime number of dollars in its bank account; in which case, if it had \$16 it would give away \$3. This might be unusual, but we can't call it irrational. An agent might prefer a dented 1973 Ford Pinto to a shiny new Mercedes. The agent might prefer prime numbers of dollars only when it owns the Pinto, but when it owns the Mercedes, it might prefer more dollars to fewer. Fortunately, the preferences of real agents are usually more systematic and thus easier to deal with.

[3] In this sense, utilities resemble temperatures: a temperature in Fahrenheit is 1.8 times the Celsius temperature plus 32, but converting from one to the other doesn't make you hotter or colder.

15.3.1 Utility assessment and utility scales

If we want to build a decision-theoretic system that helps a human make decisions or acts on his or her behalf, we must first work out what the human's utility function is. This process, often called **preference elicitation**, involves presenting choices to the human and using the observed preferences to pin down the underlying utility function.

Preference elicitation

Equation (15.2) says that there is no absolute scale for utilities, but it is helpful, nonetheless, to establish *some* scale on which utilities can be recorded and compared for any particular problem. A scale can be established by fixing the utilities of any two particular outcomes, just as we fix a temperature scale by fixing the freezing point and boiling point of water. Typically, we fix the utility of a "best possible prize" at $U(S) = u_\top$ and a "worst possible catastrophe" at $U(S) = u_\bot$. (Both of these should be finite.) **Normalized utilities** use a scale with $u_\bot = 0$ and $u_\top = 1$. With such a scale, an England fan might assign a utility of 1 to England winning the World Cup and a utility of 0 to England failing to qualify.

Normalized utilities

Given a utility scale between $u_\top$ and $u_\bot$, we can assess the utility of any particular prize S by asking the agent to choose between S and a **standard lottery** $[p, u_\top; (1-p), u_\bot]$. The probability p is adjusted until the agent is indifferent between S and the standard lottery. Assuming normalized utilities, the utility of S is given by p. Once this is done for each prize, the utilities for all lotteries involving those prizes are determined. Suppose, for example, we want to know how much our England fan values the outcome of England reaching the semi-final and then losing. We compare that outcome to a standard lottery with probability p of winning the trophy and probability $1-p$ of an ignominious failure to qualify. If there is indifference at $p = 0.3$, then 0.3 is the value of reaching the semi-final and then losing.

Standard lottery

In medical, transportation, environmental and other decision problems, people's lives are at stake. (Yes, there are things more important than England's fortunes in the World Cup.) In such cases, $u_\bot$ is the value assigned to immediate death (or in the really worst cases, many deaths). *Although nobody feels comfortable with putting a value on human life, it is a fact that tradeoffs on matters of life and death are made all the time.* Aircraft are given a complete overhaul at intervals, rather than after every trip. Cars are manufactured in a way that trades off costs against accident survival rates. We tolerate a level of air pollution that kills four million people a year.

Paradoxically, a refusal to put a monetary value on life can mean that life is *undervalued*. Ross Shachter describes a government agency that commissioned a study on removing asbestos from schools. The decision analysts performing the study assumed a particular dollar value for the life of a school-age child, and argued that the rational choice under that assumption was to remove the asbestos. The agency, morally outraged at the idea of setting the value of a life, rejected the report out of hand. It then decided against asbestos removal—implicitly asserting a lower value for the life of a child than that assigned by the analysts.

Currently several agencies of the U.S. government, including the Environmental Protection Agency, the Food and Drug Administration, and the Department of Transportation, use the **value of a statistical life** to determine the costs and benefits of regulations and interventions. Typical values in 2019 are roughly $10 million.

Value of a statistical life

Some attempts have been made to find out the value that people place on their own lives. One common "currency" used in medical and safety analysis is the **micromort**, a one in a million chance of death. If you ask people how much they would pay to avoid a risk—for

Micromort

example, to avoid playing Russian roulette with a million-barreled revolver—they will respond with very large numbers, perhaps tens of thousands of dollars, but their actual behavior reflects a much lower monetary value for a micromort.

For example, in the UK, driving in a car for 230 miles incurs a risk of one micromort. Over the life of your car—say, 92,000 miles—that's 400 micromorts. People appear to be willing to pay about $12,000 more for a safer car that halves the risk of death. Thus, their car-buying action says they have a value of $60 per micromort. A number of studies have confirmed a figure in this range across many individuals and risk types. However, government agencies such as the U.S. Department of Transportation typically set a lower figure; they will spend only about $6 in road repairs per expected life saved. Of course, these calculations hold only for small risks. Most people won't agree to kill themselves, even for $60 million.

QALY

Another measure is the **QALY**, or quality-adjusted life year. Patients are willing to accept a shorter life expectancy to avoid a disability. For example, kidney patients on average are indifferent between living two years on dialysis and one year at full health.

15.3.2 The utility of money

Utility theory has its roots in economics, and economics provides one obvious candidate for a utility measure: money (or more specifically, an agent's total net assets). The almost universal exchangeability of money for all kinds of goods and services suggests that money plays a significant role in human utility functions.

Monotonic preference

It will usually be the case that an agent prefers more money to less, all other things being equal. We say that the agent exhibits a **monotonic preference** for more money. This does not mean that money behaves as a utility function, because it says nothing about preferences between *lotteries* involving money.

Suppose you have triumphed over the other competitors in a television game show. The host now offers you a choice: either you can take the $1,000,000 prize or you can gamble it on the flip of a coin. If the coin comes up heads, you end up with nothing, but if it comes up tails, you get $2,500,000. If you're like most people, you would decline the gamble and pocket the million. Are you being irrational?

Expected monetary value

Assuming the coin is fair, the **expected monetary value** (EMV) of the gamble is $\frac{1}{2}(\$0) + \frac{1}{2}(\$2{,}500{,}000) = \$1{,}250{,}000$, which is more than the original $1,000,000. But that does not necessarily mean that accepting the gamble is a better decision. Suppose we use S_n to denote the state of possessing total wealth $\$n$, and that your current wealth is $\$k$. Then the expected utilities of the two actions of accepting and declining the gamble are

$$\begin{aligned} EU(Accept) &= \tfrac{1}{2}U(S_k) + \tfrac{1}{2}U(S_{k+2{,}500{,}000})\,, \\ EU(Decline) &= U(S_{k+1{,}000{,}000})\,. \end{aligned}$$

To determine what to do, we need to assign utilities to the outcome states. Utility is not directly proportional to monetary value, because the utility for your first million is very high (or so they say), whereas the utility for an additional million is smaller. Suppose you assign a utility of 5 to your current financial status (S_k), a 9 to the state $S_{k+2{,}500{,}000}$, and an 8 to the state $S_{k+1{,}000{,}000}$. Then the rational action would be to decline, because the expected utility of accepting is only 7 (less than the 8 for declining). On the other hand, a billionaire would most likely have a utility function that is locally linear over the range of a few million more, and thus would accept the gamble.

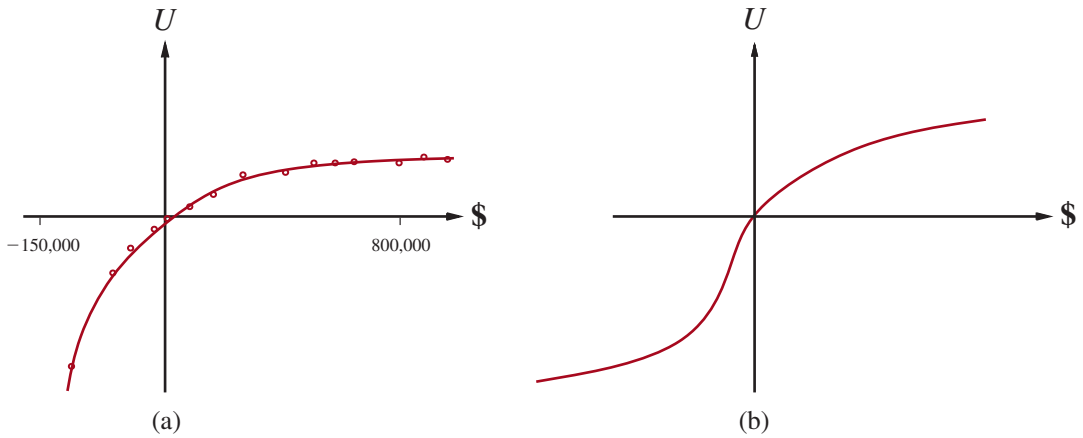


Figure 15.2 The utility of money. (a) Empirical data for Mr. Beard over a limited range. (b) A typical curve for the full range.

In a pioneering study of actual utility functions, Grayson (1960) found that the utility of money was almost exactly proportional to the *logarithm* of the amount. (This idea was first suggested by Bernoulli (1738); see Exercise 15.STPT.) One particular utility curve, for a certain Mr. Beard, is shown in Figure 15.2(a). The data obtained for Mr. Beard's preferences are consistent with a utility function

$$U(S_{k+n}) = -263.31 + 22.09 \log(n + 150,000)$$

for the range between $n = -\$150,000$ and $n = \$800,000$.

We should not assume that this is the definitive utility function for monetary value, but it is likely that most people have a utility function that is concave for positive wealth. Going into debt is bad, but preferences between different levels of debt can display a reversal of the concavity associated with positive wealth. For example, someone already \$10,000,000 in debt might well accept a gamble on a fair coin with a gain of \$10,000,000 for heads and a loss of \$20,000,000 for tails.[4] This yields the S-shaped curve shown in Figure 15.2(b).

If we restrict our attention to the positive part of the curves, where the slope is decreasing, then for any lottery L, the utility of being faced with that lottery is less than the utility of being handed the expected monetary value of the lottery as a sure thing:

$$U(L) < U(S_{EMV(L)}) .$$

That is, agents with curves of this shape are **risk-averse**: they prefer a sure thing with a payoff that is less than the expected monetary value of a gamble. On the other hand, in the "desperate" region at large negative wealth in Figure 15.2(b), the behavior is **risk-seeking**. The value an agent will accept in lieu of a lottery is called the **certainty equivalent** of the lottery. Studies have shown that most people will accept about \$400 in lieu of a gamble that gives \$1000 half the time and \$0 the other half—that is, the certainty equivalent of the lottery is \$400, while the EMV is \$500.

Risk-averse

Risk-seeking

Certainty equivalent

The difference between the EMV of a lottery and its certainty equivalent is called the **insurance premium**. Risk aversion is the basis for the insurance industry, because it means that

Insurance premium

[4] Such behavior might be called desperate, but it is rational if one is already in a desperate situation.

insurance premiums are positive. People would rather pay a small insurance premium than gamble the price of their house against the chance of a fire. From the insurance company's point of view, the price of the house is very small compared with the firm's total reserves. This means that the insurer's utility curve is approximately linear over such a small region, and the gamble costs the company almost nothing.

Risk-neutral

Notice that for *small* changes in wealth relative to the current wealth, almost any curve will be approximately linear. An agent that has a linear curve is said to be **risk-neutral**. For gambles with small sums, therefore, we expect risk neutrality. In a sense, this justifies the simplified procedure that proposed small gambles to assess probabilities and to justify the axioms of probability in Section 12.2.3.

15.3.3 Expected utility and post-decision disappointment

The rational way to choose the best action, a^*, is to maximize expected utility:

$$a^* = \operatorname*{argmax}_a EU(a)\,.$$

If we have calculated the expected utility correctly according to our probability model, and if the probability model correctly reflects the underlying stochastic processes that generate the outcomes, then, on average, we will get the utility we expect if the whole process is repeated many times.

In reality, however, our model usually oversimplifies the real situation, either because we don't know enough (e.g., when making a complex investment decision) or because the computation of the true expected utility is too difficult (e.g., when making a move in backgammon, needing to take into account all possible future dice rolls). In that case, we are really working with *estimates* $\widehat{EU}(a)$ of the true expected utility. We will assume, kindly perhaps, that the estimates are **unbiased**—that is, the expected value of the error, $E(\widehat{EU}(a) - EU(a))$, is zero. In that case, it still seems reasonable to choose the action with the highest estimated utility and to expect to receive that utility, on average, when the action is executed.

Unbiased

Unfortunately, the real outcome will usually be significantly *worse* than we estimated, even though the estimate was unbiased! To see why, consider a decision problem in which there are k choices, each of which has true estimated utility of 0. Suppose that the error in each utility estimate is independent and has a unit normal distribution—that is, a Gaussian with zero mean and standard deviation of 1, shown as the bold curve in Figure 15.3. Now, as we actually start to generate the estimates, some of the errors will be negative (pessimistic) and some will be positive (optimistic). Because we select the action with the *highest* utility estimate, we are favoring the overly optimistic estimates, and that is the source of the bias.

Order statistic

It is a straightforward matter to calculate the distribution of the maximum of the k estimates and hence quantify the extent of our disappointment. (This calculation is a special case of computing an **order statistic**, the distribution of any particular ranked element of a sample.) Suppose that each estimate X_i has a probability density function $f(x)$ and cumulative distribution $F(x)$. (As explained in Appendix A, the cumulative distribution F measures the probability that the cost is less than or equal to any given amount—that is, it integrates the original density f.) Now let X^* be the largest estimate, i.e., $\max\{X_1,\ldots,X_k\}$. Then the cumulative distribution for X^* is

$$\begin{aligned} P(\max\{X_1,\ldots,X_k\} \le x) &= P(X_1 \le x,\ldots,X_k \le x) \\ &= P(X_1 \le x)\ldots P(X_k \le x) = F(x)^k\,. \end{aligned}$$

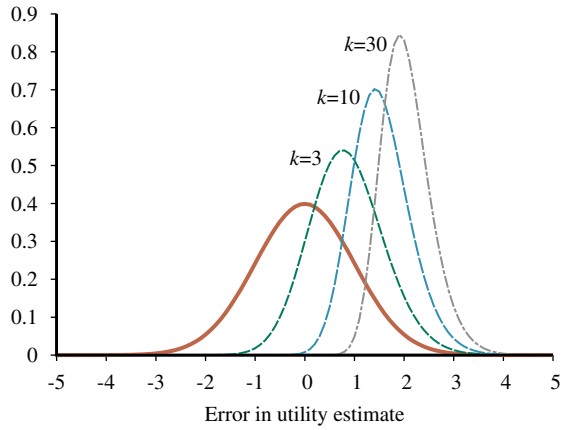


Figure 15.3 Unjustified optimism caused by choosing the best of k options: we assume that each option has a true utility of 0 but a utility estimate that is distributed according to a unit normal (brown curve). The other curves show the distributions of the maximum of k estimates for $k=3$, 10, and 30.

The probability density function is the derivative of the cumulative distribution function, so the density for X^*, the maximum of k estimates, is

$$P(x) = \frac{d}{dx}\left(F(x)^k\right) = kf(x)(F(x))^{k-1}\,.$$

These densities are shown for different values of k in Figure 15.3 for the case where $f(x)$ is the unit normal. For $k=3$, the density for X^* has a mean around 0.85, so the average disappointment will be about 85% of the standard deviation in the utility estimates. With more choices, extremely optimistic estimates are more likely to arise: for $k=30$, the disappointment will be around twice the standard deviation in the estimates.

This tendency for the estimated expected utility of the best choice to be too high is called the **optimizer's curse** (Smith and Winkler, 2006). It afflicts even the most seasoned decision analysts and statisticians. Serious manifestations include believing that an exciting new drug that has cured 80% of patients in a trial will cure 80% of patients (it's been chosen from $k=$ thousands of candidate drugs) or that a mutual fund advertised as having above-average returns will continue to have them (it's been chosen to appear in the advertisement out of $k=$ dozens of funds in the company's overall portfolio). It can even be the case that what appears to be the best choice may not be, if the variance in the utility estimate is high: a drug that has cured 9 of 10 patients and has been selected from thousands tried is probably *worse* than one that has cured 800 of 1000.

Optimizer's curse

The optimizer's curse crops up everywhere because of the ubiquity of utility-maximizing selection processes, so taking the utility estimates at face value is a bad idea. We can avoid the curse with a Bayesian approach that uses an explicit probability model $\mathbf{P}(\widehat{EU}\,|\,EU)$ of the error in the utility estimates. Given this model and a prior on what we might reasonably expect the utilities to be, we treat the utility estimate as evidence and compute the posterior distribution for the true utility using Bayes' rule.

15.3.4 Human judgment and irrationality

Normative theory
Descriptive theory

Decision theory is a **normative theory**: it describes how a rational agent *should* act. A **descriptive theory**, on the other hand, describes how actual agents—for example, humans—really do act. The application of economic theory would be greatly enhanced if the two coincided, but there appears to be some experimental evidence to the contrary. The evidence suggests that humans are "predictably irrational" (Ariely, 2009).

The best-known problem is the Allais paradox (Allais, 1953). People are given a choice between lotteries A and B and then between C and D, which have the following prizes:

A : 80% chance of \$4000 C : 20% chance of \$4000
B : 100% chance of \$3000 D : 25% chance of \$3000

Most people consistently prefer B over A (taking the sure thing), and C over D (taking the higher EMV). The normative analysis disagrees! We can see this most easily if we use the freedom implied by Equation (15.2) to set $U(\$0) = 0$. In that case, then $B \succ A$ implies that $U(\$3000) > 0.8\,U(\$4000)$, whereas $C \succ D$ implies exactly the reverse. In other words, there is no utility function that is consistent with these choices.

Certainty effect

One explanation for the apparently irrational preferences is the **certainty effect** (Kahneman and Tversky, 1979): people are strongly attracted to gains that are certain. There are several reasons why this may be so.

First, people may prefer to reduce their computational burden; by choosing certain outcomes, they don't have to compute with probabilities. But the effect persists even when the computations involved are very easy ones.

Second, people may distrust the legitimacy of the stated probabilities. I trust that a coin flip is roughly 50/50 if I have control over the coin and the flip, but I may distrust the result if the flip is done by someone with a vested interest in the outcome.[5] In the presence of distrust, it might be better to go for the sure thing.[6]

Third, people may be accounting for their emotional state as well as their financial state. People know they would experience regret if they gave up a certain reward (B) for an 80% chance at a higher reward and then lost.

In other words, if A is chosen, there is a 20% chance of getting no money *and feeling like a complete idiot*, which is worse than just getting no money. So perhaps people who choose B over A and C over D are not irrational; they are willing to give up \$200 of EMV to avoid a 20% chance of feeling like an idiot.

A related problem is the Ellsberg paradox. Here the prizes are fixed, but the probabilities are underconstrained. Your payoff will depend on the color of a ball chosen from an urn. You are told that the urn contains 1/3 red balls, and 2/3 either black or yellow balls, but you don't know how many black and how many yellow. Again, you are asked whether you prefer lottery A or B; and then C or D:

A : \$100 for a red ball C : \$100 for a red or yellow ball
B : \$100 for a black ball D : \$100 for a black or yellow ball.

It should be clear that if you think there are more red than black balls then you should prefer

[5] For example, the mathematician/magician Persi Diaconis can make a coin flip come out the way he wants every time (Landhuis, 2004).

[6] Even the sure thing may not be certain. Despite cast-iron promises, we have not yet received that \$27,000,000 from the Nigerian bank account of a previously unknown deceased relative.

A over B and C over D; if you think there are fewer red than black you should prefer the opposite. But it turns out that most people prefer A over B and also prefer D over C, even though there is no state of the world for which this is rational. It seems that people have **ambiguity aversion**: A gives you a 1/3 chance of winning, while B could be anywhere between 0 and 2/3. Similarly, D gives you a 2/3 chance, while C could be anywhere between 1/3 and 3/3. Most people elect the known probability rather than the unknown unknowns.

Ambiguity aversion

Yet another problem is that the exact wording of a decision problem can have a big impact on the agent's choices; this is called the **framing effect**. Experiments show that people like a medical procedure that is described as having a "90% survival rate" about twice as much as one described as having a "10% death rate," even though these two statements mean exactly the same thing. This discrepancy in judgment has been found in multiple experiments and is about the same whether the subjects are patients in a clinic, statistically sophisticated business school students, or experienced doctors.

Framing effect

People feel more comfortable making *relative* utility judgments rather than absolute ones. I may have little idea how much I might enjoy the various wines offered by a restaurant. The restaurant takes advantage of this by offering a $200 bottle that nobody will buy, but which serves to skew upward the customer's estimate of the value of all wines, making a $55 bottle seem like a bargain. This is called the **anchoring effect**.

Anchoring effect

If human informants insist on contradictory preference judgments, there is nothing that automated agents can do to be consistent with them. Fortunately, preference judgments made by humans are often open to revision in the light of further consideration. Paradoxes like the Allais and Ellsberg paradoxes are greatly reduced (but not eliminated) if the choices are explained better. In work at the Harvard Business School on assessing the utility of money, Keeney and Raiffa (1976, p. 210) found the following:

> Subjects tend to be too risk-averse in the small and therefore ... the fitted utility functions exhibit unacceptably large risk premiums for lotteries with a large spread. ... Most of the subjects, however, can reconcile their inconsistencies and feel that they have learned an important lesson about how they want to behave. As a consequence, some subjects cancel their automobile collision insurance and take out more term insurance on their lives.

The evidence for human irrationality is also questioned by researchers in the field of **evolutionary psychology**, who point to the fact that our brain's decision-making mechanisms did not evolve to solve word problems with probabilities and prizes stated as decimal numbers. Let us grant, for the sake of argument, that the brain has built-in neural mechanisms for computing with probabilities and utilities, or something functionally equivalent. If so, the required inputs would be obtained through accumulated experience of outcomes and rewards rather than through linguistic presentations of numerical values.

Evolutionary psychology

It is far from obvious that we can directly access the brain's built-in neural mechanisms by presenting decision problems in linguistic/numerical form. The very fact that different wordings of the *same decision problem* elicit different choices suggests that the decision problem itself is not getting through. Spurred by this observation, psychologists have tried presenting problems in uncertain reasoning and decision making in "evolutionarily appropriate" forms; for example, instead of saying "90% survival rate," the experimenter might show 100 stick-figure animations of the operation, where the patient dies in 10 of them and survives in 90. With decision problems posed in this way, people's behavior seems to be much closer to the standard of rationality.

15.4 Multiattribute Utility Functions

Decision making in the field of public policy involves high stakes, in both money and lives. For example, in deciding what levels of harmful emissions to allow from a power plant, policy makers must weigh the prevention of death and disability against the benefit of the power and the economic burden of mitigating the emissions. Picking a site for a new airport requires consideration of the disruption caused by construction; the cost of land; the distance from centers of population; the noise of flight operations; safety issues arising from local topography and weather conditions; and so on. Problems like these, in which outcomes are characterized by two or more attributes, are handled by **multiattribute utility theory**. In essence, it's the theory of comparing apples to oranges.

Multiattribute utility theory

Let the attributes be $\mathbf{X} = X_1, \ldots, X_n$ and let $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ be a complete vector of assignments, where each x_i is either a numeric value or a discrete value with an assumed ordering on values. The analysis is easier if we arrange it so that higher values of an attribute always correspond to higher utilities: utilities are monotonically increasing. That means that we can't use, say, the number of deaths, d as an attribute; we would have to use $-d$. It also means that we can't use the room temperature, t, as an attribute. If the utility function for temperature has a peak at 70°F and falls off monotonically on either side, then we could split the attribute into two pieces. We could use $t - 70$ to measure whether the room is warm enough, and $70 - t$ to measure whether it is cool enough; both of these attributes would be monotonically increasing until they reach their maximum utility value at 0; the utility curve is flat from that point on, meaning that you dont't get any more "warm enough" above 70°F, nor any more "cool enough" below 70°F.

The attributes in the airport problem could be:

- *Throughput*, measured by the number of flights per day;
- *Safety*, measured by minus the expected number of deaths per year;
- *Quietness*, measured by minus the number of people living under the flight paths;
- *Frugality*, measured by the negative cost of construction.

We begin by examining cases in which decisions can be made *without* combining the attribute values into a single utility value. Then we look at cases in which the utilities of attribute combinations can be specified very concisely.

15.4.1 Dominance

Suppose that airport site S_1 costs less, generates less noise pollution, and is safer than site S_2. One would not hesitate to reject S_2. We then say that there is **strict dominance** of S_1 over S_2. In general, if an option is of lower value on all attributes than some other option, it need not be considered further. Strict dominance is often very useful in narrowing down the field of choices to the real contenders, although it seldom yields a unique choice. Figure 15.4(a) shows a schematic diagram for the two-attribute case.

Strict dominance

That is fine for the deterministic case, in which the attribute values are known for sure. What about the general case, where the outcomes are uncertain? A direct analog of strict dominance can be constructed, where, despite the uncertainty, all possible concrete outcomes for S_1 strictly dominate all possible outcomes for S_2. (See Figure 15.4(b).) Of course, this will probably occur even less often than in the deterministic case.

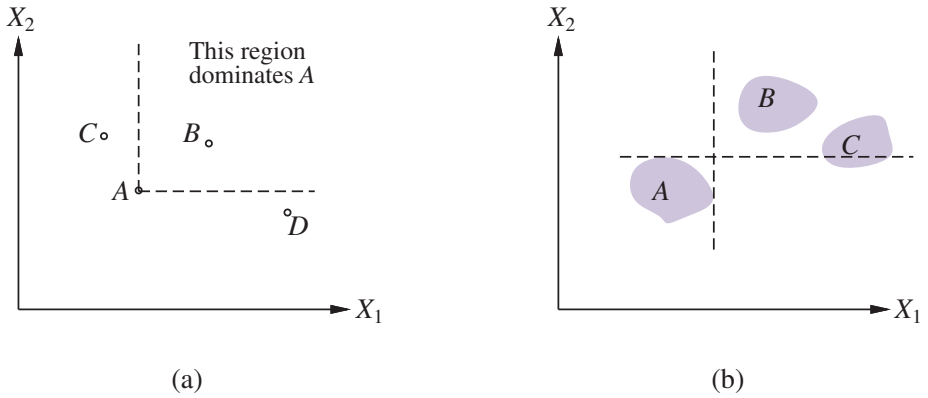


Figure 15.4 Strict dominance. (a) Deterministic: Option A is strictly dominated by B but not by C or D. (b) Uncertain: A is strictly dominated by B but not by C.

Stochastic dominance

Fortunately, there is a more useful generalization called **stochastic dominance**, which occurs very frequently in real problems. Stochastic dominance is easiest to understand in the context of a single attribute. Suppose we believe that the cost of placing the airport at S_1 is uniformly distributed between \$2.8 billion and \$4.8 billion and that the cost at S_2 is uniformly distributed between \$3 billion and \$5.2 billion. Define the *Frugality* attribute to be the negative cost. Figure 15.5(a) shows the distributions for the frugality of sites S_1 and S_2. Then, given only the information that the more frugal choice is better (all other things being equal), we can say that S_1 stochastically dominates S_2 (i.e., S_2 can be discarded). It is important to note that this does *not* follow from comparing the expected costs. For example, if we knew the cost of S_1 to be *exactly* \$3.8 billion, then we would be *unable* to make a decision without additional information on the utility of money. (It might seem odd that *more* information on the cost of S_1 could make the agent *less* able to decide. The paradox is resolved by noting that in the absence of exact cost information, the decision is easier to make but is more likely to be wrong.)

The exact relationship between the attribute distributions needed to establish stochastic dominance is best seen by examining the cumulative distributions, shown in Figure 15.5(b). If the cumulative distribution for S_1 is always to the right of the cumulative distribution for S_2, then, stochastically speaking, S_1 is cheaper than S_2. Formally, if two actions A_1 and A_2 lead to probability distributions $p_1(x)$ and $p_2(x)$ on attribute X, then A_1 stochastically dominates A_2 on X if

$$\forall x \int_{-\infty}^{x} p_1(x')\,dx' \leq \int_{-\infty}^{x} p_2(x')\,dx' .$$

The relevance of this definition to the selection of optimal decisions comes from the following property: *if A_1 stochastically dominates A_2, then for any monotonically nondecreasing utility function $U(x)$, the expected utility of A_1 is at least as high as the expected utility of A_2.* To see why this is true, consider the two expected utilities, $\int p_1(x)U(x)dx$ and $\int p_2(x)U(x)dx$. Initially, it's not obvious why the first integral is bigger than the second, given that the stochastic dominance condition has a p_1-integral that is smaller than the p_2-integral.

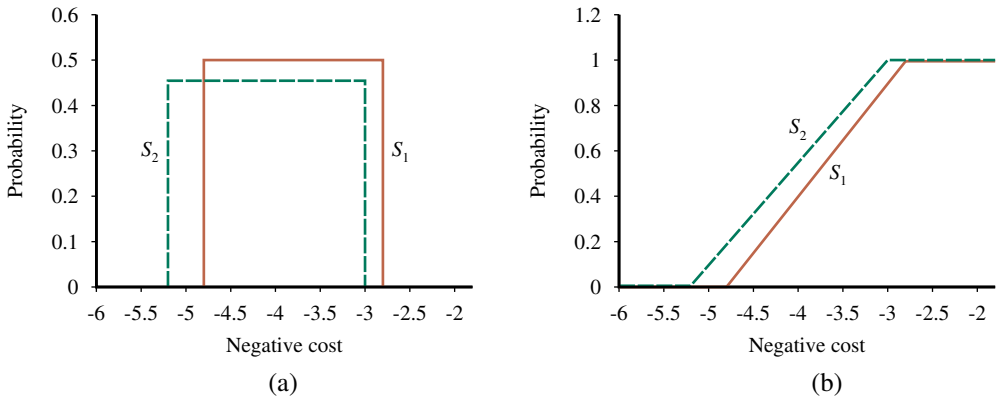


Figure 15.5 Stochastic dominance. (a) S_1 stochastically dominates S_2 on frugality (negative cost). (b) Cumulative distributions for the frugality of S_1 and S_2.

Instead of thinking about the integral over x, however, think about the integral over y, the cumulative probability, as shown in Figure 15.5(b). For any value of y, the corresponding value of x (and hence of $U(x)$) is bigger for S_1 than for S_2; so if we integrate a bigger quantity over the whole range of y, we are bound to get a bigger result. Formally, it's just a substitution of $y = P_1(x)$ in the integral for S_1's expected value and $y = P_2(x)$ in the integral for S_2's. With these substitutions, we have $dy = \frac{d}{dx}(P_1(x))dx = p_1(x)dx$ for S_1 and $dy = p_2(x)dx$ for S_2, hence

$$\int_{-\infty}^{\infty} p_1(x)U(x)dx = \int_0^1 U(P_1^{-1}(y))dy \geq \int_0^1 U(P_2^{-1}(y))dy = \int_{-\infty}^{\infty} p_2(x)U(x)dx\,.$$

This inequality allows us to prefer A_1 to A_2 in a single-attribute problem. More generally, if an action is stochastically dominated by another action on *all* attributes in a multiattribute problem, then it can be discarded.

The stochastic dominance condition might seem rather technical and perhaps not so easy to evaluate without extensive probability calculations. In fact, it can be decided very easily in many cases. For example, would you rather fall head-first onto concrete from 3 millimeters or 3 meters? Assuming you chose 3 millimeters—good choice! Why is it necessarily a better decision? There is a good deal of uncertainty about the degree of damage you will incur in both cases; but for any given level of damage, the probability that you'll incur at least that level of damage is higher when falling from 3 meters than from 3 millimeters. In other words, 3 millimeters stochastically dominates 3 meters on the *Safety* attribute.

This kind of reasoning comes as second nature to humans; it's so obvious we don't even think about it. Stochastic domination abounds in the airport problem too. Suppose, for example, that the construction transportation cost depends on the distance to the supplier. The cost itself is uncertain, but the greater the distance, the greater the cost. If S_1 is closer than S_2, then S_1 will dominate S_2 on frugality. Although we will not present them here, algorithms exist for propagating this kind of qualitative information among uncertain variables in **qualitative probabilistic networks**, enabling a system to make rational decisions based on stochastic dominance, without using any numeric values.

Qualitative probabilistic networks

15.4.2 Preference structure and multiattribute utility

Suppose we have n attributes, each of which has d distinct possible values. To specify the complete utility function $U(x_1,\ldots,x_n)$, we need d^n values in the worst case. Multiattribute utility theory aims to identify additional structure in human preferences so that we don't need to specify all d^n values individually. Having identified some regularity in preference behavior, we then derive **representation theorems** to show that an agent with a certain kind of preference structure has a utility function

Representation theorem

$$U(x_1,\ldots,x_n) = F[f_1(x_1),\ldots,f_n(x_n)]\,,$$

where F is (we hope) a simple function such as addition. Notice the similarity to the use of Bayesian networks to decompose the joint probability of several random variables.

As an example, suppose each x_i is the amount of money the agent has in a particular currency: dollars, euros, marks, lira, etc. The f_i functions could then convert each amount into a common currency, and F would then be simply addition.

Preferences without uncertainty

Let us begin with the deterministic case. On page 522 we noted that for deterministic environments, the agent has a value function, which we write here as $V(x_1,\ldots,x_n)$; the aim is to represent this function concisely. The basic regularity that arises in deterministic preference structures is called **preference independence**. Two attributes X_1 and X_2 are preferentially independent of a third attribute X_3 if the preference between outcomes $\langle x_1,x_2,x_3\rangle$ and $\langle x_1',x_2',x_3\rangle$ does not depend on the particular value x_3 for attribute X_3.

Preference independence

Going back to the airport example, where we have (among other attributes) *Quietness*, *Frugality*, and *Safety* to consider, one may propose that *Quietness* and *Frugality* are preferentially independent of *Safety*. For example, if we prefer an outcome with 20,000 people residing in the flight path and a construction cost of \$4 billion over an outcome with 70,000 people residing in the flight path and a cost of \$3.7 billion when the safety level is 0.006 deaths per billion passenger miles in both cases, then we would have the same preference when the safety level is 0.012 or 0.003; and the same independence would hold for preferences between any other pair of values for *Quietness* and *Frugality*. It is also apparent that *Frugality* and *Safety* are preferentially independent of *Quietness* and that *Quietness* and *Safety* are preferentially independent of *Frugality*.

We say that the set of attributes $\{Quietness, Frugality, Safety\}$ exhibits **mutual preferential independence (MPI)**. MPI says that, whereas each attribute may be important, it does not affect the way in which one trades off the other attributes against each other.

Mutual preferential independence (MPI)

Mutual preferential independence is a complicated name, but it leads to a simple form for the agent's value function (Debreu, 1960): *If attributes $X_1,\ldots,X_n$ are mutually preferentially independent, then the agent's preferences can be represented by a value function*

$$V(x_1,\ldots,x_n) = \sum_i V_i(x_i)\,,$$

where each V_i refers only to the attribute X_i. For example, it might well be the case that the airport decision can be made using a value function

$$V(quietness, frugality, safety) = quietness \times 10^4 + frugality + safety \times 10^{12}\,.$$

A value function of this type is called an **additive value function**. Additive functions are an

Additive value function

extremely natural way to describe an agent's preferences and are valid in many real-world situations. For n attributes, assessing an additive value function requires assessing n separate one-dimensional value functions rather than one n-dimensional function; typically, this represents an exponential reduction in the number of preference experiments that are needed. Even when MPI does not strictly hold, as might be the case at extreme values of the attributes, an additive value function might still provide a good approximation to the agent's preferences. This is especially true when the violations of MPI occur in portions of the attribute ranges that are unlikely to occur in practice.

To understand MPI better, it helps to look at cases where it *doesn't* hold. Suppose you are at a medieval market, considering the purchase of some hunting dogs, some chickens, and some wicker cages for the chickens. The hunting dogs are very valuable, but if you don't have enough cages for the chickens, the dogs will eat the chickens; hence, the tradeoff between dogs and chickens depends strongly on the number of cages, and MPI is violated. The existence of these kinds of interactions among various attributes makes it much harder to assess the overall value function.

Preferences with uncertainty

When uncertainty is present in the domain, we also need to consider the structure of preferences between lotteries and to understand the resulting properties of utility functions, rather than just value functions. The mathematics of this problem can become quite complicated, so we present just one of the main results to give a flavor of what can be done.

Utility independence

The basic notion of **utility independence** extends preference independence to cover lotteries: a set of attributes $\mathbf{X}$ is utility independent of a set of attributes $\mathbf{Y}$ if preferences between lotteries on the attributes in $\mathbf{X}$ are independent of the particular values of the attributes in $\mathbf{Y}$. A set of attributes is **mutually utility independent** (MUI) if each of its subsets is utility-independent of the remaining attributes. Again, it seems reasonable to propose that the airport attributes are MUI.

Mutually utility independent

Multiplicative utility function

MUI implies that the agent's behavior can be described using a **multiplicative utility function** (Keeney, 1974). The general form of a multiplicative utility function is best seen by looking at the case for three attributes. For conciseness, we use U_i to mean $U_i(x_i)$:

$$U = k_1U_1 + k_2U_2 + k_3U_3 + k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_3k_1U_3U_1 + k_1k_2k_3U_1U_2U_3\,.$$

Although this does not look very simple, it contains just three single-attribute utility functions and three constants. In general, an n-attribute problem exhibiting MUI can be modeled using n single-attribute utilities and n constants. Each of the single-attribute utility functions can be developed independently of the other attributes, and this combination will be guaranteed to generate the correct overall preferences. Additional assumptions are required to obtain a purely additive utility function.

15.5 Decision Networks

In this section, we look at a general mechanism for making rational decisions. The notation is often called an **influence diagram** (Howard and Matheson, 1984), but we will use the more descriptive term **decision network**. Decision networks combine Bayesian networks

Influence diagram

Decision network

with additional node types for actions and utilities. We use the problem of picking an airport site as an example.

15.5.1 Representing a decision problem with a decision network

In its most general form, a decision network represents information about the agent's current state, its possible actions, the state that will result from the agent's action, and the utility of that state. It therefore provides a substrate for implementing utility-based agents of the type first introduced in Section 2.4.5. Figure 15.6 shows a decision network for the airport-siting problem. It illustrates the three types of nodes used:

- **Chance nodes** (ovals) represent random variables, just as they do in Bayesian networks. The agent could be uncertain about the construction cost, the level of air traffic and the potential for litigation, and the *Safety*, *Quietness*, and total *Frugality* variables, each of which also depends on the site chosen. Each chance node has associated with it a conditional distribution that is indexed by the state of the parent nodes. In decision networks, the parent nodes can include decision nodes as well as chance nodes. Note that each of the current-state chance nodes could be part of a large Bayesian network for assessing construction costs, air traffic levels, or litigation potentials. Chance nodes
- **Decision nodes** (rectangles) represent points where the decision maker has a choice of actions. In this case, the *AirportSite* action can take on a different value for each site under consideration. The choice influences the safety, quietness, and frugality of the solution. In this chapter, we assume that we are dealing with a single decision node. Chapter 16 deals with cases in which more than one decision must be made. Decision nodes
- **Utility nodes** (diamonds) represent the agent's utility function.[7] The utility node has as parents all variables describing the outcomes that directly affect utility. Associated with the utility node is a description of the agent's utility as a function of the parent attributes. The description could be just a tabulation of the function, or it might be a parameterized additive or linear function of the attribute values. For now, we will assume that the function is deterministic; that is, given the values of its parent variables, the value of the utility node is fully determined. Utility nodes

A simplified form is also used in many cases. The notation remains identical, but the chance nodes describing the outcome states are omitted. Instead, the utility node is connected directly to the current-state nodes and the decision node. In this case, rather than representing a utility function on outcome states, the utility node represents the *expected* utility associated with each action, as defined in Equation (15.1) on page 519; that is, the node is associated with an **action-utility function** (also known as a **Q-function** in reinforcement learning, as described in Chapter 23). Figure 15.7 shows the action-utility representation of the airport siting problem. Action-utility function

Notice that, because the *Quietness*, *Safety*, and *Frugality* chance nodes in Figure 15.6 refer to future states, they can never have their values set as evidence variables. Thus, the simplified version that omits these nodes can be used whenever the more general form can be used. Although the simplified form contains fewer nodes, the omission of an explicit description of the outcome of the siting decision means that it is less flexible with respect to changes in circumstances.

[7] These nodes are also called **value nodes** in the literature.

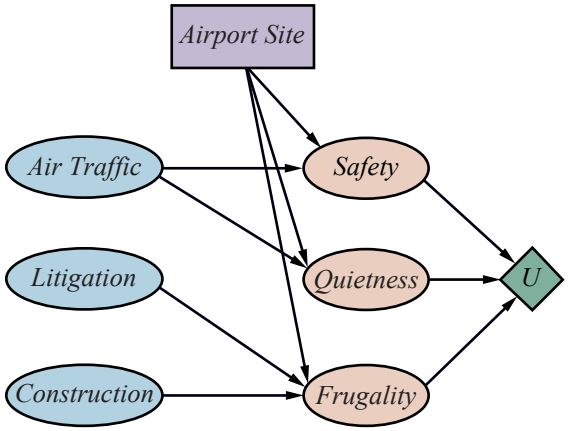


Figure 15.6 A decision network for the airport-siting problem.

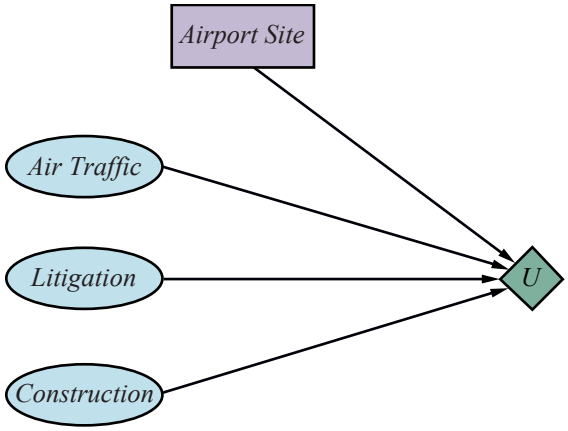


Figure 15.7 A simplified representation of the airport-siting problem. Chance nodes corresponding to outcome states have been factored out.

For example, in Figure 15.6, a change in aircraft noise levels can be reflected by a change in the conditional probability table associated with the *Quietness* node, whereas a change in the weight accorded to noise pollution in the utility function can be reflected by a change in the utility table. In the action-utility diagram, Figure 15.7, on the other hand, all such changes have to be reflected by changes to the action-utility table. Essentially, the action-utility formulation is a *compiled* version of the original formulation, obtained by summing out the outcome state variables.

15.5.2 Evaluating decision networks

Actions are selected by evaluating the decision network for each possible setting of the decision node. Once the decision node is set, it behaves exactly like a chance node that has been set as an evidence variable. The algorithm for evaluating decision networks is the following:

1. Set the evidence variables for the current state.
2. For each possible value of the decision node:
   (a) Set the decision node to that value.
   (b) Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm.
   (c) Calculate the resulting utility for the action.
3. Return the action with the highest utility.

This is a straightforward approach that can utilize any available Bayesian network algorithm and can be incorporated directly into the agent design given in Figure 12.1 on page 406. We will see in Chapter 16 that the possibility of executing several actions in sequence makes the problem much more interesting.

15.6 The Value of Information

In the preceding analysis, we have assumed that all relevant information, or at least all available information, is provided to the agent before it makes its decision. In practice, this is hardly ever the case. *One of the most important parts of decision making is knowing what questions to ask.* For example, a doctor cannot expect to be provided with the results of all possible diagnostic tests and questions at the time a patient first enters the consulting room. Tests are often expensive and sometimes hazardous (both directly and because of associated delays). Their importance depends on two factors: whether the test results would lead to a significantly better treatment plan, and how likely the various test results are.

Information value theory

This section describes **information value theory**, which enables an agent to choose what information to acquire. We assume that prior to selecting a "real" action represented by the decision node, the agent can acquire the value of any of the potentially observable chance variables in the model. Thus, information value theory involves a simplified form of sequential decision making—simplified because the observation actions affect only the agent's **belief state**, not the external physical state. The value of any particular observation must derive from the potential to affect the agent's eventual physical action; and this potential can be estimated directly from the decision model itself.

15.6.1 A simple example

Suppose an oil company is hoping to buy one of n indistinguishable blocks of ocean-drilling rights. Let us assume further that exactly one of the blocks contains oil that will generate net profits of C dollars, while the others are worthless. The asking price of each block is C/n dollars. If the company is risk-neutral, then it will be indifferent between buying a block and not buying one because the expected profit is zero in both cases.

Now suppose that a seismologist offers the company the results of a survey of block number 3, which indicates definitively whether the block contains oil. How much should the company be willing to pay for the information? The way to answer this question is to examine what the company would do if it had the information:

- With probability $1/n$, the survey will indicate oil in block 3. In this case, the company will buy block 3 for C/n dollars and make a profit of $C - C/n = (n-1)C/n$ dollars.

- With probability $(n-1)/n$, the survey will show that the block contains no oil, in which case the company will buy a different block. Now the probability of finding oil in one of the other blocks changes from $1/n$ to $1/(n-1)$, so the company makes an expected profit of $C/(n-1) - C/n = C/n(n-1)$ dollars.

Now we can calculate the expected profit, given access to the survey information:

$$\frac{1}{n} \times \frac{(n-1)C}{n} + \frac{n-1}{n} \times \frac{C}{n(n-1)} = C/n\,.$$

Thus, the information is worth C/n dollars to the company, and the company should be willing to pay the seismologist some significant fraction of this amount.

The value of information derives from the fact that *with* the information, one's course of action can be changed to suit the *actual* situation. One can discriminate according to the situation, whereas without the information, one has to do what's best on average over the possible situations. In general, the value of a given piece of information is defined to be the difference in expected value between best actions before and after information is obtained.

15.6.2 A general formula for perfect information

Value of perfect information

It is simple to derive a general mathematical formula for the value of information. We assume that exact evidence can be obtained about the value of some random variable E_j (that is, we learn $E_j = e_j$), so the phrase **value of perfect information** (VPI) is used.[8]

In the agent's initial information state, the value of the current best action α is, from Equation (15.1),

$$EU(\alpha) = \max_a \sum_{s'} P(\text{RESULT}(a) = s')\, U(s')\,,$$

and the value of the new best action (after the new evidence $E_j = e_j$ is obtained) will be

$$EU(\alpha_{e_j} | e_j) = \max_a \sum_{s'} P(\text{RESULT}(a) = s' \,|\, e_j)\, U(s')\,.$$

But E_j is a random variable whose value is *currently* unknown, so to determine the value of discovering E_j we must average over all possible values e_j that we might discover for E_j, using our *current* beliefs about its value:

$$VPI(E_j) = \left(\sum_{e_j} P(E_j = e_j)\, EU(\alpha_{e_j} | E_j = e_j) \right) - EU(\alpha)\,.$$

To get some intuition for this formula, consider the simple case where there are only two actions, a_1 and a_2, from which to choose. Their current expected utilities are U_1 and U_2. The information $E_j = e_j$ will yield some new expected utilities U_1' and U_2' for the actions, but before we obtain E_j, we will have some probability distributions over the possible values of U_1' and U_2' (which we assume are independent).

Suppose that a_1 and a_2 represent two different routes through a mountain range in winter: a_1 is a nice, straight highway through a tunnel, and a_2 is a winding dirt road over the top. Just

[8] There is no loss of expressiveness in requiring perfect information. Suppose we wanted to model the case in which we become somewhat more certain about a variable. We can do that by introducing *another* variable about which we learn perfect information. For example, suppose we initially have broad uncertainty about the variable *Temperature*. Then we gain the perfect knowledge *Thermometer* = 37; this gives us imperfect information about the true *Temperature*, and the uncertainty due to measurement error is encoded in the sensor model $\mathbf{P}(\textit{Thermometer} \,|\, \textit{Temperature})$. See Exercise 15.VPIX for another example.

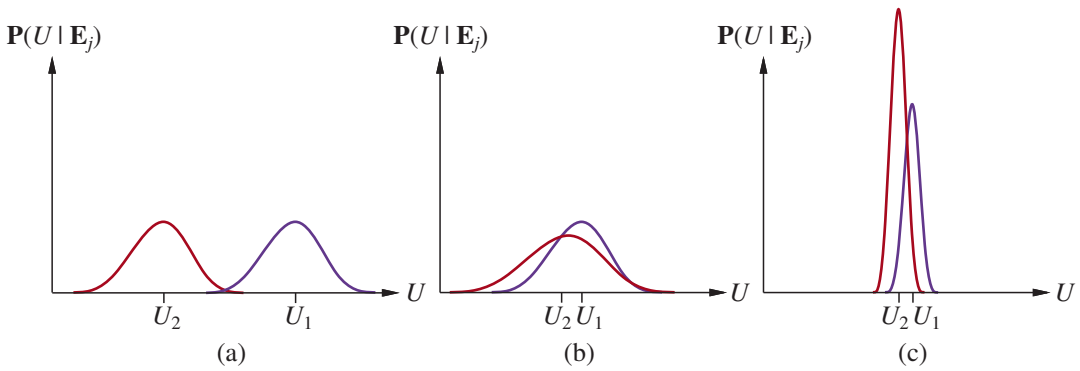


Figure 15.8 Three generic cases for the value of information. In (a), a_1 will almost certainly remain superior to a_2, so the information is not needed. In (b), the choice is unclear and the information is crucial. In (c), the choice is unclear, but because it makes little difference, the information is less valuable. (Note: The fact that U_2 has a high peak in (c) means that its expected value is known with higher certainty than U_1.)

given this information, a_1 is clearly preferable, because it is quite possible that a_2 is blocked by snow, whereas it is unlikely that anything blocks a_1. U_1 is therefore clearly higher than U_2. It is possible to obtain satellite reports E_j on the actual state of each road that would give new expectations, U_1' and U_2', for the two crossings. The distributions for these expectations are shown in Figure 15.8(a). Obviously, in this case, it is not worth the expense of obtaining satellite reports, because it is unlikely that the information derived from them will change the plan. With no change, information has no value.

Now suppose that we are choosing between two different winding dirt roads of slightly different lengths and we are carrying a seriously injured passenger. Then, even when U_1 and U_2 are quite close, the distributions of U_1' and U_2' are very broad. There is a significant possibility that the second route will turn out to be clear while the first is blocked, and in this case the difference in utilities will be very high. The VPI formula indicates that it might be worthwhile getting the satellite reports. Such a situation is shown in Figure 15.8(b).

Finally, suppose that we are choosing between the two dirt roads in summertime, when blockage by snow is unlikely. In this case, satellite reports might show one route to be more scenic than the other because of flowering alpine meadows, or perhaps wetter because of recent rain. It is therefore quite likely that we would change our plan if we had the information. In this case, however, the difference in value between the two routes is still likely to be very small, so we will not bother to obtain the reports. This situation is shown in Figure 15.8(c).

In sum, *information has value to the extent that it is likely to cause a change of plan and to the extent that the new plan will be significantly better than the old plan.*

15.6.3 Properties of the value of information

One might ask whether it is possible for information to be deleterious: can it actually have negative expected value? Intuitively, one should expect this to be impossible. After all, one could in the worst case just ignore the information and pretend that one has never received it. This is confirmed by the following theorem, which applies to any decision-theoretic agent using any decision network with possible observations E_j:

The expected value of information is nonnegative:

$$\forall\, j \;\; VPI(E_j) \geq 0\,.$$

The theorem follows directly from the definition of VPI, and we leave the proof as an exercise (Exercise 15.NNVP). It is, of course, a theorem about *expected* value, not *actual* value. Additional information can easily lead to a plan that *turns out to* be worse than the original plan if the information happens to be misleading. For example, a medical test that gives a false positive result may lead to unnecessary surgery; but that does not mean that the test shouldn't be done.

It is important to remember that VPI depends on the current state of information. It can change as more information is acquired. For any given piece of evidence E_j, the value of acquiring it can go down (e.g., if another variable strongly constrains the posterior for E_j) or up (e.g., if another variable provides a clue on which E_j builds, enabling a new and better plan to be devised). Thus, VPI is not additive. That is,

$$VPI(E_j, E_k) \neq VPI(E_j) + VPI(E_k) \qquad \text{(in general)}\,.$$

VPI is, however, order-independent. That is,

$$VPI(E_j, E_k) = VPI(E_j) + VPI(E_k|E_j) = VPI(E_k) + VPI(E_j|E_k) = VPI(E_k, E_j)$$

where the notation $VPI(\cdot|E)$ denotes the VPI calculated according to the posterior distribution where E is already observed. Order independence distinguishes sensing actions from ordinary actions and simplifies the problem of calculating the value of a sequence of sensing actions. We return to this question in the next section.

15.6.4 Implementation of an information-gathering agent

A sensible agent should ask questions in a reasonable order, should avoid asking questions that are irrelevant, should take into account the importance of each piece of information in relation to its cost, and should stop asking questions when that is appropriate. All of these capabilities can be achieved by using the value of information as a guide.

Figure 15.9 shows the overall design of an agent that can gather information intelligently before acting. For now, we assume that with each observable evidence variable E_j, there is an associated cost, $C(E_j)$, which reflects the cost of obtaining the evidence through tests, consultants, questions, or whatever. The agent requests what appears to be the most efficient observation in terms of utility gain per unit cost. We assume that the result of the action $Request(E_j)$ is that the next percept provides the value of E_j. If no observation is worth its cost, the agent selects a "real" action.

Myopic

The agent algorithm we have described implements a form of information gathering that is called **myopic**. This is because it uses the VPI formula shortsightedly, calculating the value of information as if only a single evidence variable will be acquired. Myopic control is based on the same heuristic idea as greedy search and often works well in practice. (For example, it has been shown to outperform expert physicians in selecting diagnostic tests.) However, if there is no single evidence variable that will help a lot, a myopic agent might hastily take an action when it would have been better to request two or more variables first and then take action. The next section considers the possibility of obtaining multiple observations.

```
function INFORMATION-GATHERING-AGENT(percept) returns an action
  persistent: D, a decision network

  integrate percept into D
  j ← the value that maximizes VPI(E_j) / C(E_j)
  if VPI(E_j) > C(E_j)
      then return Request(E_j)
  else return the best action from D
```

Figure 15.9 Design of a simple, myopic information-gathering agent. The agent works by repeatedly selecting the observation with the highest information value, until the cost of the next observation is greater than its expected benefit.

15.6.5 Nonmyopic information gathering

The fact that the value of a sequence of observations is invariant under permutations of the sequence is intriguing but doesn't, by itself, lead to efficient algorithms for optimal information gathering. Even if we restrict ourselves to choosing in advance a fixed subset of observations to collect, there are 2^n possible such subsets from n potential observations. In the general case, we face an even more complex problem of finding an optimal *conditional plan* (as described in Section 11.5.2) that chooses an observation and then acts or chooses more observations, depending on the outcome. Such plans form trees, and the number of such trees is superexponential in n.[9]

For observations of variables in a decision network, it turns out that this problem is intractable even when the network is a polytree. There are, however, special cases in which the problem can be solved efficiently. Here we present one such case: the **treasure hunt** problem (or the **least-cost testing sequence** problem, for the less romantically inclined). There are n locations $1,\ldots,n$; each location i contains treasure with independent probability $P(i)$; and it costs $C(i)$ to check location i. This corresponds to a decision network where all the potential evidence variables $Treasure_i$ are absolutely independent. The agent examines locations in some order until treasure is found; the question is, what is the optimal order?

Treasure hunt

To answer this question, we will need to consider the expected costs and success probabilities of various sequences of observations, assuming the agent stops when treasure is found. Let $\mathbf{x}$ be such a sequence; $\mathbf{xy}$ be the concatenation of sequences $\mathbf{x}$ and $\mathbf{y}$; $C(\mathbf{x})$ be the expected cost of $\mathbf{x}$; $P(\mathbf{x})$ be the probability that sequence $\mathbf{x}$ succeeds in finding treasure; and $F(\mathbf{x}) = 1 - P(\mathbf{x})$ be the probability that it fails. Given these definitions, we have

$$C(\mathbf{xy}) = C(\mathbf{x}) + F(\mathbf{x})C(\mathbf{y}), \tag{15.3}$$

that is, the sequence $\mathbf{xy}$ will definitely incur the cost of $\mathbf{x}$ and, if $\mathbf{x}$ fails, it will also incur the cost of $\mathbf{y}$.

The basic idea in any sequence optimization problem is to look at the change in cost, defined by $\Delta = C(\mathbf{wxyz}) - C(\mathbf{wyxz})$, when two adjacent subsequences $\mathbf{x}$ and $\mathbf{y}$ in a general sequence $\mathbf{wxyz}$ are flipped. When the sequence is optimal, all such changes make the sequence worse. The first step is to show that the sign of the effect (increasing or decreasing

[9] The general problem of generating sequential behavior in a partially observable environment falls under the heading of **partially observable Markov decision processes**, which are described in Chapter 16.

the cost) doesn't depend on the context provided by $\mathbf{w}$ and $\mathbf{z}$. We have

$$\begin{aligned}\Delta &= [C(\mathbf{w}) + F(\mathbf{w})C(\mathbf{xyz})] - [C(\mathbf{w}) + F(\mathbf{w})C(\mathbf{yxz})] \qquad \text{(by Equation (15.3))} \\ &= F(\mathbf{w})[C(\mathbf{xyz}) - C(\mathbf{yxz})] \\ &= F(\mathbf{w})[(C(\mathbf{xy}) + F(\mathbf{xy})C(\mathbf{z})) - (C(\mathbf{yx}) + F(\mathbf{yx})C(\mathbf{z}))] \quad \text{(by Equation (15.3))} \\ &= F(\mathbf{w})[C(\mathbf{xy}) - C(\mathbf{yx})] \qquad \text{(since } F(\mathbf{xy}) = F(\mathbf{yx})\text{)}.\end{aligned}$$

So we have shown that the direction of the change in the cost of the whole sequence depends only on the direction of the change in cost of the pair of elements being flipped; the context of the pair doesn't matter. This gives us a way to sort the sequence by pairwise comparisons to obtain an optimal solution. Specifically, we now have

$$\begin{aligned}\Delta &= F(\mathbf{w})[(C(\mathbf{x}) + F(\mathbf{x})C(\mathbf{y})) - (C(\mathbf{y}) + F(\mathbf{y})C(\mathbf{x}))] \qquad \text{(by Equation (15.3))} \\ &= F(\mathbf{w})[C(\mathbf{x})(1 - F(\mathbf{y})) - C(\mathbf{y})(1 - F(\mathbf{x}))] = F(\mathbf{w})[C(\mathbf{x})P(\mathbf{y}) - C(\mathbf{y})P(\mathbf{x})]\,.\end{aligned}$$

This holds for any sequences $\mathbf{x}$ and $\mathbf{y}$, so it holds specifically when $\mathbf{x}$ and $\mathbf{y}$ are single observations of locations i and j, respectively. So we know that, for i and j to be adjacent in an optimal sequence, we must have $C(i)P(j) \leq C(j)P(i)$, or $\frac{P(i)}{C(i)} \geq \frac{P(j)}{C(j)}$. In other words, the optimal order ranks the locations according to the success probability per unit cost. Exercise 15.HUNT asks you to determine whether this is in fact the policy followed by the algorithm in Figure 15.9 for this problem.

15.6.6 Sensitivity analysis and robust decisions

Sensitivity analysis

The practice of **sensitivity analysis** is widespread in technological disciplines: it means analyzing how much the output of a process changes as the model parameters are tweaked. Sensitivity analysis in probabilistic and decision-theoretic systems is particularly important because the probabilities used are typically either learned from data or estimated by human experts, which means that they are themselves subject to considerable uncertainty. Only in rare cases, such as the dice rolls in backgammon, are the probabilities objectively known.

For a utility-driven decision-making process, you can think of the output as either the actual decision made or the expected utility of that decision. Consider the latter first: because expectation depends on probabilities from the model, we can compute the derivative of the expected utility of any given action with respect to each of those probability values. (For example, if all the conditional probability distributions in the model are explicitly tabulated, then computing the expectation involves computing a ratio of two sum-of-product expressions; for more on this, see Chapter 21.) Thus, one can determine which parameters in the model have the largest effect on the expected utility of the final decision.

If, instead, we are concerned about the actual decision made, rather than its utility according to the model, then we can simply vary the parameters systematically (perhaps using binary search) to see whether the decision changes, and, if so, what is the smallest perturbation that causes such a change. One might think it doesn't matter that much which decision is made, only what its utility is. That's true, but in practice there may be a very substantial difference between the *real* utility of a decision and the utility *according to the model*.

If all reasonable perturbations of the parameters leave the optimal decision unchanged, then it is reasonable to assume the decision is a good one, even if the utility estimate for that decision is substantially incorrect. If, on the other hand, the optimal decision changes considerably as the parameters of the model change, then there is a good chance that the

model may produce a decision that is substantially suboptimal in reality. In that case, it is worth investing further effort to refine the model.

These intuitions have been formalized in several fields (control theory, decision analysis, risk management) that propose the notion of a **robust** or **minimax** decision—that is, one that gives the best result in the worst case. Here, "worst case" means worst with respect to all plausible variations in the parameter values of the model. Letting θ stand for all the parameters in the model, the robust decision is defined by

Robust

$$a^* = \operatorname*{argmax}_a \min_\theta EU(a;\theta) \ .$$

In many cases, particularly in control theory, the robust approach leads to designs that work very reliably in practice. In other cases, it leads to overly conservative decisions. For example, when designing a self-driving car, the robust approach would assume the worst case for the behavior of the other vehicles on the road—that is, they are all driven by homicidal maniacs. In that case, the optimal solution for the car is to stay in the garage.

Bayesian decision theory offers an alternative to robust methods: if there is uncertainty about the parameters of the model, then model that uncertainty using hyperparameters.

Whereas the robust approach might say that some probability θ_i in the model could be anywhere between 0.3 and 0.7, with the actual value chosen by an adversary to make things come out as badly as possible, the Bayesian approach would put a prior probability distribution on θ_i and then proceed as before. This requires more modeling effort—for example, the Bayesian modeler must decide if parameters θ_i and θ_j are independent—but often results in better performance in practice.

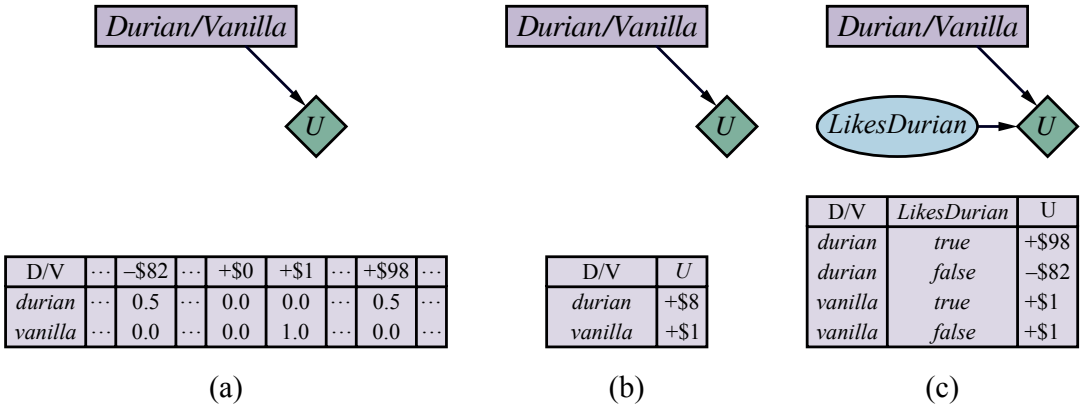
In addition to parametric uncertainty, applications of decision theory in the real world also suffer from *structural* uncertainty. For example, the assumption of independence of *AirTraffic*, *Litigation*, and *Construction* in Figure 15.6 may be incorrect, and there may be additional variables that the model simply omits. At present, we do not have a good understanding of how to take this kind of uncertainty into account. One possibility is to keep an ensemble of models, perhaps generated by machine learning algorithms, in the hope that the ensemble captures the significant variations that matter.

15.7 Unknown Preferences

In this section we discuss what happens when there is uncertainty about the utility function whose expected value is to be optimized. There are two versions of this problem: one in which an agent (machine or human) is uncertain about its *own* utility function, and another in which a machine is supposed to help a human but is uncertain about what the human wants.

15.7.1 Uncertainty about one's own preferences

Imagine that you are at an ice-cream shop in Thailand and they have only two flavors left: vanilla and durian. Both cost \$2. You know you have a moderate liking for vanilla and you'd be willing to pay up to \$3 for a vanilla ice cream on such a hot day, so there is a net gain of \$1 for choosing vanilla. On the other hand, you have no idea whether you like durian or not, but you've read on Wikipedia that the durian elicits different responses from different people: some find that "it surpasses in flavour all other fruits of the world" while others liken it to "sewage, stale vomit, skunk spray and used surgical swabs."



D/V	⋯	–$82	⋯	+$0	+$1	⋯	+$98	⋯
durian	⋯	0.5	⋯	0.0	0.0	⋯	0.5	⋯
vanilla	⋯	0.0	⋯	0.0	1.0	⋯	0.0	⋯

(a)

D/V	*U*
durian	+$8
vanilla	+$1

(b)

D/V	*LikesDurian*	U
durian	*true*	+$98
durian	*false*	–$82
vanilla	*true*	+$1
vanilla	*false*	+$1

(c)

Figure 15.10 (a) A decision network for the ice cream choice with an uncertain utility function. (b) The network with the expected utility of each action. (c) Moving the uncertainty from the utility function into a new random variable.

To put some numbers on this, let's say there's a 50% chance you'll find it sublime (+$100) and a 50% chance you'll hate it (-$80 if the taste lingers all afternoon). Here, there's no uncertainty about what prize you're going to win—it's the same durian ice cream either way—but there's uncertainty about your own preferences for that prize.

We could extend the decision network formalism to allow for uncertain utilities, as shown in Figure 15.10(a). If there is no more information to be obtained about your durian preferences, however—for example, if the shop won't let you taste it first—then the decision problem is identical to the one shown in Figure 15.10(b). We can simply replace the uncertain value of the durian with its expected net gain of $(0.5 \times \$100) - (0.5 \times \$80) - \$2 = \$8$ and your decision will remain unchanged.

If it's possible for your beliefs about durian to change—perhaps you get a tiny taste, or you find out that all of your living relatives love durian—then the transformation in Figure 15.10(b) is not valid. It turns out, however, that we can still find an equivalent model in which the utility function is deterministic. Rather than saying there is uncertainty about the utility function, we move that uncertainty "into the world," so to speak. That is, we create a new random variable *LikesDurian* with prior probabilities of 0.5 for *true* and *false*, as shown in Figure 15.10(c). With this extra variable, the utility function becomes deterministic, but we can still handle changing beliefs about your durian preferences.

The fact that unknown preferences can be modeled by ordinary random variables means that we can keep using the machinery and theorems developed for known preferences. On the other hand, it doesn't mean that we can always assume that preferences are known. The uncertainty is still there and still affects how agents should behave.

15.7.2 Deference to humans

Now let's turn to the second case mentioned above: a machine that is supposed to help a human but is uncertain about what the human wants. The full treatment of this case must be deferred to Chapter 17, where we discuss decisions involving more than one agent. Here, we ask one simple question: under what circumstances will such a machine defer to the human?

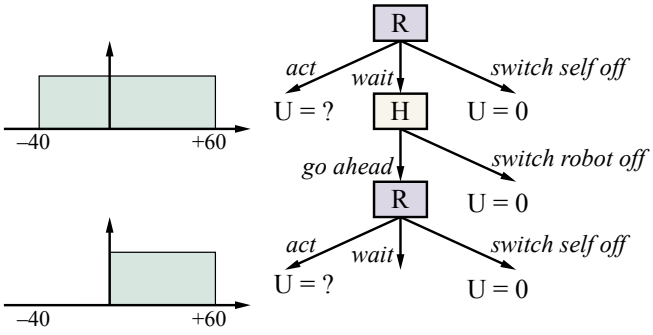


Figure 15.11 The off-switch game. *R*, the robot, can choose to act now, with a highly uncertain payoff; to switch itself off; or to defer to *H*, the human. *H* can switch *R* off or let it go ahead. *R* now has the same choice again. Acting still has an uncertain payoff, but now *R* knows the payoff is nonnegative.

To study this question, let's consider a very simple scenario, as shown in Figure 15.11. Robbie is a software robot working for Harriet, a busy human, as her personal assistant. Harriet needs a hotel room for her next business meeting in Geneva. Robbie can act now—let's say he can book Harriet into a very expensive hotel near the meeting venue. He is quite unsure how much Harriet will like the hotel and its price; let's say he has a uniform probability for its net value to Harriet between -40 and $+60$, with an average of $+10$. He could also "switch himself off"—less melodramatically, take himself out of the hotel booking process altogether—which we define (without loss of generality) to have value 0 to Harriet. If those were his two choices, he would go ahead and book the hotel, incurring a significant risk of making Harriet unhappy. (If the range were -60 to $+40$, with average -10, he would switch himself off instead.) We'll give Robbie a third choice, however: explain his plan, wait, and let Harriet switch him off. Harriet can either switch him off or let him go ahead and book the hotel. What possible good could this do, one might ask, given that he could make both of those choices himself?

The point is that Harriet's choice—to switch Robbie off or let him go ahead—provides Robbie with information about Harriet's preferences. We'll assume, for now, that Harriet is rational, so if Harriet lets Robbie go ahead, it means the value to Harriet is positive. Now, as shown in Figure 15.11, Robbie's belief changes: it is uniform between 0 and $+60$, with an average of $+30$.

So, if we evaluate Robbie's initial choices from his point of view:

1. Acting now and booking the hotel has an expected value of $+10$.
2. Switching himself off has a value of 0.
3. Waiting and letting Harriet switch him off leads to two possible outcomes:
   (a) There is a 40% chance, based on Robbie's uncertainty about Harriet's preferences, that she will hate the plan and will switch Robbie off, with value 0.
   (b) There is a 60% chance Harriet will like the plan and allow Robbie to go ahead, with expected value $+30$.

Thus, waiting has expected value $(0.4 \times 0) + (0.6 \times 30) = +18$, which is better than the $+10$ Robbie expects if he acts now.

The upshot is that Robbie has a positive incentive to defer to Harriet—that is, to allow himself to be switched off. This incentive comes directly from Robbie's uncertainty about Harriet's preferences. Robbie is aware that there's a chance (40% in this example) that he might be about to do something that will make Harriet unhappy, in which case being switched off would be preferable to going ahead. Were Robbie already certain about Harriet's preferences, he would just go ahead and make the decision (or switch himself off); there would be absolutely nothing to be gained from consulting Harriet, because, according to Robbie's definite beliefs, he can already predict exactly what she is going to decide.

In fact, it is possible to prove the same result in the general case: as long as Robbie is not completely certain that he's about to do what Harriet herself would do, he is better off allowing her to switch him off. Intuitively, her decision provides Robbie with information, and the expected value of information is always nonnegative. Conversely, if Robbie is certain about Harriet's decision, her decision provides no new information, and so Robbie has no incentive to allow her to decide.

Formally, let $P(u)$ be Robbie's prior probability density over Harriet's utility for the proposed action a. Then the value of going ahead with a is

$$EU(a) = \int_{-\infty}^{\infty} P(u) \cdot u \, du = \int_{-\infty}^{0} P(u) \cdot u \, du + \int_{0}^{\infty} P(u) \cdot u \, du .$$

(We will see shortly why the integral is split up in this way.) On the other hand, the value of action d, deferring to Harriet, is composed of two parts: if $u > 0$ then Harriet lets Robbie go ahead, so the value is u, but if $u < 0$ then Harriet switches Robbie off, so the value is 0:

$$EU(d) = \int_{-\infty}^{0} P(u) \cdot 0 \, du + \int_{0}^{\infty} P(u) \cdot u \, du .$$

Comparing the expressions for $EU(a)$ and $EU(d)$, we see immediately that

$$EU(d) \geq EU(a)$$

because the expression for $EU(d)$ has the negative-utility region zeroed out. The two choices have equal value only when the negative region has zero probability—that is, when Robbie is already certain that Harriet likes the proposed action.

There are some obvious elaborations on the model that are worth exploring immediately. The first elaboration is to impose a cost for Harriet's time. In that case, Robbie is less inclined to bother Harriet if the downside risk is small. This is as it should be. And if Harriet is really grumpy about being interrupted, she shouldn't be too surprised if Robbie occasionally does things she doesn't like.

The second elaboration is to allow for some probability of human error—that is, Harriet might sometimes switch Robbie off even when his proposed action is reasonable, and she might sometimes let Robbie go ahead even when his proposed action is undesirable. It is straightforward to fold this error probability into the model (see Exercise 15.OFFS). As one might expect, the solution shows that Robbie is less inclined to defer to an irrational Harriet who sometimes acts against her own best interests. The more randomly she behaves, the more uncertain Robbie has to be about her preferences before deferring to her. Again, this is as it should be: for example, if Robbie is a self-driving car and Harriet is his naughty two-year-old passenger, Robbie should not allow Harriet to switch him off in the middle of the highway.

Summary

This chapter shows how to combine utility theory with probability to enable an agent to select actions that will maximize its expected performance.

- **Probability theory** describes what an agent should believe on the basis of evidence, **utility theory** describes what an agent wants, and **decision theory** puts the two together to describe what an agent should do.
- We can use decision theory to build a system that makes decisions by considering all possible actions and choosing the one that leads to the best expected outcome. Such a system is known as a **rational agent**.
- Utility theory shows that an agent whose preferences between lotteries are consistent with a set of simple axioms can be described as possessing a utility function; furthermore, the agent selects actions as if maximizing its expected utility.
- **Multiattribute utility theory** deals with utilities that depend on several distinct attributes of states. **Stochastic dominance** is a particularly useful technique for making unambiguous decisions, even without precise utility values for attributes.
- **Decision networks** provide a simple formalism for expressing and solving decision problems. They are a natural extension of Bayesian networks, containing decision and utility nodes in addition to chance nodes.
- Sometimes, solving a problem involves finding more information before making a decision. The **value of information** is defined as the expected improvement in utility compared with making a decision without the information; it is particularly useful for guiding the process of information-gathering prior to making a final decision.
- When, as is often the case, it is impossible to specify the human's utility function completely and correctly, machines must operate under uncertainty about the true objective. This makes a significant difference when the possibility exists for the machine to acquire more information about human preferences. We showed by a simple argument that uncertainty about preferences ensures that the machine defers to the human, to the point of allowing itself to be switched off.

Bibliographical and Historical Notes

In the 17th century treatise *L'art de Penser*, or *Port-Royal Logic*, Arnauld (1662) states:

> To judge what one must do to obtain a good or avoid an evil, it is necessary to consider not only the good and the evil in itself, but also the probability that it happens or does not happen; and to view geometrically the proportion that all these things have together.

Modern texts talk of *utility* rather than good and evil, but this statement correctly notes that one should multiply utility by probability ("view geometrically") to give expected utility, and maximize that over all outcomes ("all these things") to "judge what one must do." It is remarkable how much Arnauld got right, more than 350 years ago, and only 8 years after Pascal and Fermat first showed how to use probability correctly.

Daniel Bernoulli (1738), investigating the St. Petersburg paradox (see Exercise 15.STPT), was the first to realize the importance of preference measurement for lotteries, writing "the

value of an item must not be based on its *price*, but rather on the *utility* that it yields" (italics his). Utilitarian philosopher Jeremy Bentham (1823) proposed the **hedonic calculus** for weighing "pleasures" and "pains," arguing that all decisions (not just monetary ones) could be reduced to utility comparisons.

Hedonic calculus

Bernoulli's introduction of utility—an internal, subjective quantity—to explain human behavior via a mathematical theory was an utterly remarkable proposal for its time. It was all the more remarkable for the fact that unlike monetary amounts, the utility values of various bets and prizes are not directly observable; instead, utilities are to be inferred from the preferences exhibited by an individual. It would be two centuries before the implications of the idea were fully worked out and it became broadly accepted by statisticians and economists.

The derivation of numerical utilities from preferences was first carried out by Ramsey (1931); the axioms for preference in the present text are closer in form to those rediscovered in *Theory of Games and Economic Behavior* (von Neumann and Morgenstern, 1944). Ramsey had derived subjective probabilities (not just utilities) from an agent's preferences; Savage (1954) and Jeffrey (1983) carry out more recent constructions of this kind. Beardon *et al.* (2002) show that a utility function does not suffice to represent nontransitive preferences and other anomalous situations.

In the post-war period, decision theory became a standard tool in economics, finance, and management science. A field of **decision analysis** emerged to aid in making policy decisions more rational in areas such as military strategy, medical diagnosis, public health, engineering design, and resource management. The process involves a **decision maker** who states preferences between outcomes and a **decision analyst** who enumerates the possible actions and outcomes and elicits preferences from the decision maker to determine the best course of action. Von Winterfeldt and Edwards (1986) provide a nuanced perspective on decision analysis and its relationship to human preference structures. Smith (1988) gives an overview of the methodology of decision analysis.

Decision analysis

Decision maker

Decision analyst

Until the 1980s, multivariate decision problems were handled by constructing "decision trees" of all possible instantiations of the variables. Influence diagrams or decision networks, which take advantage of the same conditional independence properties as Bayesian networks, were introduced by Howard and Matheson (1984), based on earlier work at SRI (Miller *et al.*, 1976). Howard and Matheson's algorithm constructed the complete (exponentially large) decision tree from the decision network. Shachter (1986) developed a method for making decisions based directly on a decision network, without the creation of an intermediate decision tree. This algorithm was also one of the first to provide complete inference for multiply connected Bayesian networks. Nilsson and Lauritzen (2000) link algorithms for decision networks to ongoing developments in clustering algorithms for Bayesian networks. The collection by Oliver and Smith (1990) has a number of useful early articles on decision networks, as does the 1990 special issue of the journal *Networks*. The text by Fenton and Neil (2018) provides a hands-on guide to solving real-world decision problems using decision networks. Papers on decision networks and utility modeling also appear regularly in the journals *Management Science* and *Decision Analysis*.

Surprisingly few early AI researchers adopted decision-theoretic tools after the early applications in medical decision making described in Chapter 12. One of the few exceptions was Jerry Feldman, who applied decision theory to problems in vision (Feldman and Yakimovsky, 1974) and planning (Feldman and Sproull, 1977). Rule-based expert systems of the

late 1970s and early 1980s concentrated on answering questions, rather than on making decisions. Those systems that did recommend actions generally did so using condition–action rules rather than explicit representations of outcomes and preferences.

Decision networks offer a far more flexible approach, for example by allowing preferences to change while keeping the transition model constant, or vice versa. They also allow a principled calculation of what information to seek next. In the late 1980s, partly due to Pearl's work on Bayes nets, decision-theoretic expert systems gained widespread acceptance (Horvitz *et al.*, 1988; Cowell *et al.*, 2002). In fact, from 1991 onward, the cover design of the journal *Artificial Intelligence* has depicted a decision network, although some artistic license appears to have been taken with the direction of the arrows.

Practical attempts to measure human utilities began with post-war decision analysis (see above). The micromort utility measure is discussed by Howard (1989). Thaler Thaler (1992) found that for a 1/1000 chance of death, a respondent wouldn't pay more than $200 to remove the risk, but wouldn't accept $50,000 to take on the risk.

The use of **QALY**s (quality-adjusted life years) to perform cost–benefit analyses of medical interventions and related social policies dates back at least to work by Klarman *et al.* (1968), although the term itself was first used by Zeckhauser and Shepard (1976). Like money, QALYs correspond directly to utilities only under fairly strong assumptions, such as risk neutrality, that are often violated (Beresniak *et al.*, 2015); nonetheless, QALYs are much widely used in practice, for example in forming National Health Service policies in the UK. See Russell (1990) for a typical example of an argument for a major change in public health policy on grounds of increased expected utility measured in QALYs.

Keeney and Raiffa (1976) give an introduction to **multiattribute utility theory**. They describe early computer implementations of methods for eliciting the necessary parameters for a multiattribute utility function and include extensive accounts of real applications of the theory. Abbas (2018) covers many advances since 1976. The theory was introduced to AI primarily by the work of Wellman (1985), who also investigated the use of stochastic dominance and qualitative probability models (Wellman, 1988, 1990a). Wellman and Doyle (1992) provide a preliminary sketch of how a complex set of utility-independence relationships might be used to provide a structured model of a utility function, in much the same way that Bayesian networks provide a structured model of joint probability distributions. Bacchus and Grove (1995, 1996) and La Mura and Shoham (1999) give further results along these lines. Boutilier *et al.* (2004) describe CP-nets, a fully worked out graphical model formalism for conditional *ceteribus paribus* preference statements.

The **optimizer's curse** was brought to the attention of decision analysts in a forceful way by Smith and Winkler (2006), who pointed out that the financial benefits to the client projected by analysts for their proposed course of action almost never materialized. They trace this directly to the bias introduced by selecting an optimal action and show that a more complete Bayesian analysis eliminates the problem.

Post-decision disappointment

The same underlying concept has been called **post-decision disappointment** by Harrison and March (1984) and was noted in the context of analyzing capital investment projects by Brown (1974). The optimizer's curse is also closely related to the **winner's curse** (Capen *et al.*, 1971; Thaler, 1992), which applies to competitive bidding in auctions: whoever wins the auction is very likely to have overestimated the value of the object in question. Capen *et al*. quote a petroleum engineer on the topic of bidding for oil-drilling rights: "If one wins a

Winner's curse

tract against two or three others he may feel fine about his good fortune. But how should he feel if he won against 50 others? Ill."

The Allais paradox, due to Nobel Prize–winning economist Maurice Allais (1953), was tested experimentally to show that people are consistently inconsistent in their judgments (Tversky and Kahneman, 1982; Conlisk, 1989). The Ellsberg paradox on ambiguity aversion was introduced in the Ph.D. thesis of Daniel Ellsberg (1962).[10] Fox and Tversky (1995) describe a further study of ambiguity aversion. Machina (2005) gives an overview of choice under uncertainty and how it can vary from expected utility theory. See the classic text by Keeney and Raiffa (1976) and the more recent work by Abbas (2018) for an in-depth analysis of preferences with uncertainty.

Irrationality

2009 was a big year for popular books on human **irrationality**, including *Predictably Irrational* (Ariely, 2009), *Sway* (Brafman and Brafman, 2009), *Nudge* (Thaler and Sunstein, 2009), *Kluge* (Marcus, 2009), *How We Decide* (Lehrer, 2009) and *On Being Certain* (Burton, 2009). They complement the classic book *Judgment Under Uncertainty* (Kahneman *et al.*, 1982) and the article that started it all (Kahneman and Tversky, 1979). Kahneman himself provides an insightful and readable account in *Thinking: Fast and Slow* (Kahneman, 2011).

The field of evolutionary psychology (Buss, 2005), on the other hand, has run counter to this literature, arguing that humans are quite rational in evolutionarily appropriate contexts. Its adherents point out that irrationality is penalized by definition in an evolutionary context and show that in some cases it is an artifact of the experimental setup (Cummins and Allen, 1998). There has been a recent resurgence of interest in Bayesian models of cognition, overturning decades of pessimism (Elio, 2002; Chater and Oaksford, 2008; Griffiths *et al.*, 2008); this resurgence is not without its detractors, however (Jones and Love, 2011).

The theory of information value was explored first in the context of statistical experiments, where a quasi-utility (entropy reduction) was used (Lindley, 1956). The control theorist Ruslan Stratonovich (1965) developed the more general theory presented here, in which information has value by virtue of its ability to affect decisions. Stratonovich's work was not known in the West, where Ron Howard (1966) pioneered the same idea. His paper ends with the remark "If information value theory and associated decision theoretic structures do not in the future occupy a large part of the education of engineers, then the engineering profession will find that its traditional role of managing scientific and economic resources for the benefit of man has been forfeited to another profession." To date, the implied revolution in managerial methods has not occurred.

The myopic information-gathering algorithm described in the chapter is ubiquitous in the decision analysis literature; its basic outlines can be discerned in the original paper on influence diagrams (Howard and Matheson, 1984). Efficient calculation methods are studied by Dittmer and Jensen (1997). Laskey (1995) and Nielsen and Jensen (2003) discuss methods for sensitivity analysis in Bayesian networks and decision networks, respectively. The classic text *Robust and Optimal Control* (Zhou *et al.*, 1995) provides thorough coverage and comparison of the robust and decision-theoretic approaches to decisions under uncertainty.

The treasure hunt problem was solved independently by many authors, dating back at least to papers on sequential testing by Gluss (1959) and Mitten (1960). The style of proof

[10] Ellsberg later became a military analyst at the RAND Corporation and leaked documents known as the Pentagon Papers, thereby contributing to the end of the Vietnam war.

in this chapter draws on a basic result, due to Smith (1956), relating the value of a sequence to the value of the same sequence with two adjacent elements permuted. These results for independent tests were extended to more general tree and graph search problems (where the tests are partially ordered) by Kadane and Simon (1977). Results on the complexity of non-myopic calculations of the value of information were obtained by Krause and Guestrin (2009). Krause *et al.* (2008) identified cases where submodularity leads to a tractable approximation algorithm, drawing on the seminal work of Nemhauser *et al.* (1978) on submodular functions; Krause and Guestrin (2005) identify cases where an exact dynamic programming algorithm gives an efficient solution for both evidence subset election and conditional plan generation.

Harsanyi (1967) studied the problem of *incomplete* information in game theory, where players may not know each others' payoff functions exactly. He showed that such games were identical to games with *imperfect* information, where players are uncertain about the state of the world, via the trick of adding state variables referring to players' payoffs. Cyert and de Groot (1979) developed a theory of **adaptive utility** in which an agent could be uncertain about its own utility function and could obtain more information through experience.

Adaptive utility

Work on Bayesian preference elicitation (Chajewska *et al.*, 2000; Boutilier, 2002) begins from the assumption of a prior probability over the agent's utility function. Fern *et al.* (2014) propose a decision-theoretic model of **assistance** in which a robot tries to ascertain and assist with a human goal about which it is initially uncertain. The off-switch example in Section 15.7.2 is adapted from Hadfield-Menell *et al.* (2017b). Russell (2019) proposes a general framework for benefical AI in which the off-switch game is a key example.

Assistance