1

Decision Problems

1.1 maximisation of expected benefit

Decision theory begins with decision problems. Decision problems arise for *agents*: entities with the resources to represent, evaluate and change the world around them in various different ways, typically within the context of ongoing personal and institutional projects, activities or responsibilities. These projects together with the environment, both natural and social, provide the givens for the decision problems agents face: their material resources for acting, their information and often their standards for evaluating outcomes, as well as the source of the problems they must respond to. Social scientists hold very different views about the relative importance of the different aspects of this background and the decisions that are made within it, but few would doubt that choices made by consumers, doctors, policy makers and so on have the power to shape the course of events.

To face a genuine decision problem, agents must have options: actions that they are capable of performing and, equally, of forgoing if they so choose. To get an idea of what sorts of things count as a decision problem, let's look at a few examples.

1. *Take a bus?* You have an appointment that you don't want to miss. If you walk you will arrive late. If you take the bus and the traffic is light, you should arrive ahead of time. On the other hand, if the traffic is heavy then you will arrive very late, perhaps so late that the appointment will be lost. Should you risk your appointment by taking the bus?
2. *Buy health insurance?* You are currently in good health but know that if you were to fall ill you might not be able to continue to earn an income and, in the worst case, you might not be able to afford the health care

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

TABLE 1.1. *Take a Bus?*

	Heavy traffic	*Light traffic*
Take a bus	Arrive late Pay for a ticket	Arrive early Pay for a ticket
Walk	Arrive a little late No ticket needed	Arrive a little late No ticket needed

you require. By buying health insurance you can ensure that you have all the care you need. But it is 'expensive' and if your health remains good, the money is wasted. Is it worth insuring yourself?

3. *Free condoms*. By supplying free condoms, rates of transmission of venereal disease can be considerably reduced. But there is the possibility that it will also encourage sexual activity, thereby partially, or perhaps even completely, offsetting the benefits of a decreased transmission rate by virtue of the increase in the number of sexual liaisons. Should they be supplied free of charge?
4. *Vaccinations*. A vaccine has been developed for cervical cancer, a fairly common type of cancer with a high mortality rate. The vaccine is expensive, but if it is developed as part of a large-scale vaccination programme the costs are not exorbitant. The vaccine does, however, have severe side effects in very rare cases (fewer than 1 in 100,000). Should the government offer the vaccine to everyone, actively encourage them to be vaccinated or even introduce compulsory vaccination?

Decision problems such as these can be described in the following way. A decision maker has one or more options before him. The exercise of each option is associated with a number of possible consequences; some of which are desirable from the perspective of the decision maker's goals, others are not. Which consequence will result from the exercise of an option depends on the prevailing features of the environment: whether traffic is light or heavy, how much it is raining, whether you fall ill, and so on.

Let us call the set of environmental features relevant to the determination of the consequences of the exercise of any of the options a 'state of the world'. Then a decision problem can be represented by a matrix showing, for each available option, the consequence that follows from its exercise in each relevant state of the world. In our first example, for instance, taking the bus has the consequence of having to buy a ticket and arriving late in the event of heavy traffic and of paying for a ticket and arriving early in the event of light traffic. This decision problem can be represented in a simple way as in

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

TABLE 1.2. *State–Consequence Matrix*

	States			
Options	State S_1	State S_2	...	State S_n
α	A_1	A_2	...	A_n
β	B_1	B_2	...	B_n
...	...	...	...	...
γ	C_1	C_2	...	C_n

Table 1.1, where the consequences of the available options, for each of the states, are given in the table cells.

More generally, suppose that α, β, ..., and γ are the options open to the decision maker and that S_1 through S_n are n possible states of the world (these must be mutually exclusive and exhaust all the possibilities). For any option γ, let C_1 through C_n be the n consequences that might follow from exercising it. Then a decision problem can be represented by a state-consequence matrix of the kind displayed in Table 1.2.

Given a decision problem of this kind, standard decision theory says that the decision maker should choose the option whose exercise has the *greatest expected benefit*, where benefit is relative to the decision maker's evaluation of the desirability of the possible consequences of her actions. If she knows what the actual state of the world is then she should simply pick the option with the most desirable consequence in that state. Typically, however, the decision maker will be uncertain as to what the actual state is. In this case, she must consider how probable it is that each of the states is the case and pick the option whose expected desirability is greatest given these probability judgements.

For instance, suppose that I consider the probability of heavy traffic to be one-half and the benefit or utility of the various possible consequences to be as below:

	0.5	0.5
Take a bus	3	0
Walk	1	1

Then the expected benefit of taking the bus is a probability weighted average of the benefits of its possible consequences – i.e. $(3 \times 0.5) + (0 \times 0.5) = 1.5$. On the other hand, walking has a certain benefit of 1. So in this case I should take the bus. But had the probability of heavy traffic been a lot greater, walking would have been the better option.

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

TABLE 1.3. *Probability–Utility Matrix*

	Probabilities of states		
Options	$P(S_1)$	...	$P(S_n)$
α	$U(A_1)$	...	$U(A_n)$
β	$U(B_1)$	...	$U(B_n)$
...	...	...	...
γ	$U(C_1)$	...	$U(C_n)$

The next couple of chapters will be devoted to qualifying, expanding and commenting on the claim illustrated in this simple example, namely that we should pick the option that maximises expected utility. But before we do so, it will be helpful to express it more formally so that the core content is clear. Let P be a probability measure on the states of the world and U a utility measure on consequences (we will say more about what these measures are and where they come from in due course). Then a state–consequence matrix, such as that of Table 1.2, induces another matrix in which options appear as random variables: functions that assign a utility value to each state of the world (intuitively, the utility of the consequence of exercising the option in question in that state). This matrix is given in Table 1.3.

So represented, each option has an expected value that is jointly determined by the functions U and P. For instance, the expected value of option γ, denoted by $\mathbb{E}(\gamma)$, is given by $U(C_1) \cdot P(S_1) + ... + U(C_n) \cdot P(S_n)$. More generally, if the number of possible states of the world is finite:[1]

$$\mathbb{E}(\gamma) = \sum_{i=1}^{n} U(C_i).P(S_i)$$

Now what decision theory says is that rational agents should choose the option with the highest expected value. This is known as the **maximisation of expected utility hypothesis**.

The maximisation hypothesis forms the core of Bayesian decision theory, together with claims about how uncertainty should be represented and resolved through learning (respectively discussed in Chapters 3 and 10). I will argue that this hypothesis is essentially correct for cases in which we can adequately represent the decision problem we face in a manner similar to that of Table 1.2 and Table 1.3 – i.e. when we can display the problem in

[1] The restriction to a finite number of states of the world is made for simplicity, but the expected value will still be well defined even if we drop it.

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

a state–consequence matrix and can reach probability and utility judgements on all the relevant factors displayed in it. When we cannot (which is quite often the case) then the theory is not false but inapplicable, and much of the last part of this book will be devoted to answering the question as to what we do then. But for the moment our focus will be on understanding what the maximisation of expected utility hypothesis says, examining in this chapter how decision problems should framed and, in the next, how the hypothesis should be interpreted and what notion of rationality it presupposes.

1.2 FRAMING DECISIONS

Decision theory makes a claim about what option(s) it is rational to choose when the decision problem faced by the agent can be represented by a state–consequence matrix of the kind exemplified by Table 1.2. It is very important to stress that the theory does not say that you *must* frame decision problems in this way. Nor does it say that agents *will* always do so. It just says that, *if* they are framed in this way, then only options which maximise expected benefit should be chosen. Nothing precludes the possibility that the same decision situation can be framed in different ways. This is true in more than one sense.

Firstly, it may be that the problem is not naturally represented by a state–consequence decision matrix. As John Broome (1991) points out, the consequences of an action may be distributed across different times or places or people, as well as across states. The desirability of ordering a cold beer or not, for instance, will depend on the location of its consequences: it's good if the beer is served to me, in the evening, with a smile and when I have not had a few too many already; bad when it's for my children, or first thing in the morning or during a philosophy lecture. In this case, my decision problem is better represented by a matrix that associates each action and relevant combination of locations (person, time, place, etc.) with a consequence, rather than by a simple state–consequence one.

Secondly, the problem may not be representable by any kind of decision matrix at all because we are unable to identify the various elements of it: what our options are, what the relevant factors are that determine the consequence of each option, or what the consequences are of exercising one or another of the identified options when these factors are present. In particular, we may not be able to assign a determinate consequence to each state of the world for each option if the world is non-deterministic or if we cannot enumerate all the relevant conditions. This problem, of what I will term option uncertainty, is discussed in detail in Section 3.2.

Finally, it is typically possible to represent the decision problem one faces by any number of different decision matrices that differ in terms of the features of the problem that they explicitly pick out. This is true even if we just confine attention to state–consequence matrices (as I shall do), for our state of uncertainty can be more or less elaborately described by representing more or fewer of the contingencies upon which our decision might depend.

This last point raises the question of whether all such representations are equally good, or whether some are better than others. There are two claims that I want to make in this regard: firstly, that not all representations of a decision problem are equally good; and, secondly, that many representations are nonetheless permissible. This latter point is of some importance because it follows that an adequate decision theory must be ‘tolerant’ to some degree of the manner in which a problem is represented and that the solution it gives to a decision problem should be independent of the choice of representation

Let us start with the first claim, that some representations of a problem are better than others. A representation of a decision problem should help us arrive at a decision by highlighting certain features of the problem and, in particular, those upon which the decision depends. What makes one way of framing the problem better than another is simply that it is more helpful in this regard. There are at least two considerations that need to be traded off when talking about the usefulness of a representation: the expected quality of the decisions likely to be obtained and the efficiency of obtaining them. Let me say something about them both.

Quality: To make a good decision, a decision maker must give appropriate weight to the factors upon which the decision depends. In deciding whether to take an umbrella or not, for instance, I need to identify both the features of the possible outcomes of doing so that matter to me (e.g. getting wet versus staying dry) and the features of the environment upon which these outcomes depend (e.g. the eventuality of rain). Furthermore, I need to determine how significant these features are: how desirable staying dry is relative to getting wet, how probable it is that it will rain, and so on. If my representation of the decision problem is too sparse, I risk omitting features that are relevant to the decision. If I omit possible weather states from my representation of the umbrella-taking decision, for instance, then I may fail to take into account factors (in particular, the probability of rain) upon which the correctness of the decision depends. So, *ceteris paribus*, a representation that includes more relevant features will be better than one that does not.

Efficiency One way of ensuring that no relevant features are omitted is simply to list *all* the features of possible outcomes and states of the world. But drawing up and making use of such a list is clearly beyond our human capabilities and those of any real agents. Reaching judgements costs in terms of time and effort. If we try to consider all possible features of the world we will simply run out of time and energy before making a decision. A framing that delivers accuracy but is so complex that it is impossible to specify all the required inputs and to compute the expected utilities is clearly not of much use. More generally, representations that include too many features will result in inefficient decision making requiring more resources than is justified (what level of resources is justified will of course depend on what is at stake). So, *ceteris paribus*, a simpler representation will be better than a more complicated one.

Achieving a good trade-off between quality and efficiency is not just a matter of getting the level of complexity right. It is also a matter of identifying the most useful features to represent explicitly. It is useful to represent a feature if it is (sufficiently) relevant to the decision and if we can determine what significance to attach to it. A feature of the state of the world or of a consequence is relevant to a decision problem if the choice of action is sensitive to values that we might reasonably assign to this feature (its probability or utility). More precisely, one feature is more relevant than another just in case the expected values of the various actions or options under consideration are more sensitive to changes in the values of the former than the latter. For instance, whether it is desirable to take an umbrella with me or not will be sensitive to the probability of rain, but not sensitive at all to the probability of a dust storm on Mars. Likewise it is sensitive to the utility of my getting wet but not to my getting hungry, since my getting wet depends causally on the taking of the umbrella but not my getting hungry. So a good representation of my decision problem will include weather states and 'wet/dry' consequences, but not Martian dust storm states or 'hungry' consequences.

The second aspect of usefulness is equally important. A representation should be appropriate to our informational resources and our cognitive capabilities in specifying features of the environment that we are capable of tracking and features of consequences that we are capable of evaluating. If the weather is relevant to my decision as to take an umbrella or not, but I am incapable of reaching a judgement as to whether it is likely to rain or not (perhaps I have no information relevant to the question or I don't understand the information I have been given), then there is little point in framing the decision problem in terms of weather contingencies. A good representation of a problem helps us to bring the judgements we are able to make to bear on the decision problem.

It follows that whether a framing is a useful one or not will depend on properties of the decision maker (and in more than one way). Firstly, whether the features of the problem it represents are relevant depends on what matters to the decision maker and hence what sort of considerations her decisions will be sensitive to. And, secondly, whether a representation facilitates decision making will depend on the cognitive abilities and resources of the decision maker. Both of these will vary from decision maker to decision maker and from one time and context to another.

It is clearly desirable therefore that a decision theory be representation-tolerant to as great a degree as possible, in the sense of being applicable to a decision problem irrespective of how it turns out to be useful for the decision maker to represent it. Not all decision theories are equal in this regard. On the contrary, as we shall see in the next section, some impose quite severe restrictions on how a decision problem must be represented if the theory is to be used and hence make considerable demands on the decision maker in terms of the number and complexity of judgements that he must reach. Given our aim of a decision theory with a human face, this feature will count heavily against such theories.

1.3 SAVAGE'S THEORY

The modern theory of decision making under uncertainty has its roots in eighteenth-century debates over the value of gambles, with Daniel Bernoulli (1954) giving the earliest precise statement of something akin to the principle of maximising expected utility. The first axiomatic derivation of an expected utility representation of preferences is due to Frank Ramsey (1990/1926) whose treatment in many ways surpasses those of later authors. But modern decision theory descends from Savage, not Ramsey, and it is in his book *The Foundations of Statistics* that we find the first simultaneous derivation of subjective probabilities and utilities from what are clearly candidate rationality conditions on preference.

It is to Leonard Savage too that we owe the representation of decision problems faced by agents under conditions of uncertainty that was described at the beginning of the chapter and that is now standard in decision theory. Its cornerstone is a tripartite distinction between states, consequences and actions. Consequences are the features of the world that the agent cares about and seeks to bring about or avoid by acting; they are, he says, 'anything that may happen to an agent' or 'anything at all about which the person could possibly be concerned' (Leonard Savage, 1974/1954, pp. 13–14). States are those features of the world that are outside the agent's control but determine what consequence follows from the choice of action. Actions are the link between the two; formally, for Savage, they are just functions from states to consequences.

Although the distinction between states, consequences and actions is natural and useful, Savage's theory imposes some quite stringent conditions on how these objects are to be conceived. Firstly, in order that decision problems be representable by state-consequence matrices of the kind given in Table 1.2, he requires that the states of the world suffice to determine the consequence of a choice of action (I will discuss this in more detail in the next chapter). Secondly, he requires that the states themselves be causally and probabilistically independent of the action performed. And thirdly, he requires that the desirability of consequences be independent both of the state of the world in which they are realised and of the action. Jointly these assumptions imply that actions differ in value only insofar as they determine different ordered sets of consequences.

To ensure that the second two conditions hold, Savage suggests that consequences be maximally specific with regard to all that matters to the agent so that there be no uncertainty about how beneficial or desirable the consequence

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

TABLE 1.4. *Insurance Purchase*

	Good health	*Poor health*
Purchase insurance	Earn full annual income Make policy payments	Reduced income Insurance pays out
Don't purchase	Earn full annual income No policy payments	Reduced income No payout

is that derives from uncertainty about the state of the world. It follows that the states themselves must be maximally specific, 'leaving no relevant aspect undescribed' (ibid. p. 14), for if this were not the case then there could be features of the consequences of actions that matter to the agent but which are not determined by the prevailing state. So when an agent regards a Savage-style action as open to her, she must take it that she can bring it about that a maximally specific consequence will obtain conditional on each maximally specific state of the world prevailing. This is rather different from what we colloquially understand by an action. When I must choose between walking to the shops or taking the bus, as in the decision problem represented by Table 1.1, I do not do so in the light of anything like full knowledge of the consequences, in each possible state of the world, of these actions. My understanding of them is inevitably coarse-grained to some extent. It would seem, then, that Savage's theory is not well suited to agents like us, who cannot typically represent decision problems in the way required for application of his theory.

Savage was perfectly aware of this objection and drew an important distinction between small-world and grand-world decision problems. Grand-world decision problems are ones which have consequences that are maximally specific with regard to all matters of concern to the agent; small-world problems are ones with coarse-grained specifications of states and consequences. Although his theory is designed for grand-world problems, Savage argues that it could nonetheless be applied to small-world problems, so long as we ensure that the coarse-grained representation of the decision problem has sufficiently similar properties to the fine-grained one that it could be given a numerical representation by a probability–utility matrix of the kind exhibited in Table 1.3. For this, the two conditions of probabilistic independence of states from actions and desirabilistic independence of consequences from states are essential.

It is quite easy to fall foul of these constraints. Suppose that we are deciding whether to purchase health insurance for the coming years and that we represent our decision problem by the state–consequence matrix displayed in Table 1.4. So represented, it looks like a purely financial decision, that can be made on the basis of the expected incomes associated with the two

https://doi.org/10.1017/9780511760105.003 Published online by Cambridge University Press

possible acts. It is quite conceivable, however, that the value we attach to income depends on our state of health. We might need more money if our health is poor, for instance, in order to buy services that we can no longer provide for ourselves. This would be a reason to value a particular income more highly if it is gained under poor health than if it is gained under good health. Alternatively, we may get less enjoyment from money when our health is poor, and so we would value it less. Either way, in order to use Savage's theory to make a decision as to whether to purchase insurance or not, in a way that appropriately reflects the sensitivity of the desirability of money on health states, the decision problem must be reframed.

The obvious way of doing this is to take the consequences of options to be combinations of outcomes and the states in which they are realised. The act of buying health insurance, for instance, may be said to have the consequence 'Earn full income, make policy payments, enjoy good health' in good-health states and the consequence 'Earn reduced income, make policy payments, enjoy poor health' in ill-health ones. For reasons that we will examine more closely later on, Savage requires, however, that consequences and states be logically independent. So he is forced to insist that decision makers describe the consequences of their actions in terms which eliminate the sensitivity of their value to the state of the world. This is not straightforward. The way in which income varies with health states is likely to be mediated by an enormous number of variables, including the amount of support that can be expected from friends and family, the services provided by the state, charities or other institutions to help those in poor health, and one's level of psychological well-being. All of these would have to be specified in order for the act of purchasing health insurance to have a state-independent consequence in each health state. We are rarely able to do this.

A second problem in our example concerns the description of the relevant states, for the purchase of health insurance can have a causal effect on how much care one takes of one's health, so that the probability of good health is not independent of the purchase of health insurance. This problem of *moral hazard*, as it often called, plagues insurance markets. When you sell someone fire insurance, for example, you change her incentives in such a way as to make it more probable that a fire will occur. Knowing that they will be reimbursed if a fire occurs, individuals may be less careful. In extreme cases, when the value of the policy is high enough, they may even commit arson. Insurance companies have to be very careful when selling fire insurance not to underestimate their exposure. In order to eliminate the causal dependence of states on actions fuelling moral hazard the decision problem has to be reframed. In our example this would require identifying all those factors (genetic, environmental, historical) mediating the relationship between purchases of health insurance and health states, combinations of which would serve as states in the reframed decision problem. This can be very difficult to do.

The upshot is that Savage's theory is far from being representation-tolerant in the way that I argued was desirable. It is often possible to ensure that for all practical purposes any one of the three conditions required for application of his theory can be met by being careful about how the decision problem is framed. But ensuring that all three are satisfied at the same time is very difficult indeed, since the demands they impose on the description of the decision problem pull in different directions. Ensuring a determinate consequence for each state is most easily achieved by coarsening the description of outcomes, for instance, but ensuring that they have a state-independent utility requires refining them.

This problem provides strong grounds for turning our attention to a rival version of Bayesian decision theory that is due to Richard Jeffrey and Ethan Bolker. Jeffrey (1990/65) makes two modifications to the Savage framework. First, he recognises that the distinction between states and consequences is both context-and agent-dependent: that it will rain is a possible state of interest to a farmer, but a consequence for a shaman with a rain dance repertoire; that there will be flooding in low-lying areas is a possible state of the world from the perspective of a person buying a house, but a consequence from the point of view of environmental policy. So, instead of distinguishing between the objects of belief and those of desire, he takes the contents of all the decision maker's attitudes to be propositions. This small modification has a very important implication. Since events and consequences are logically interrelated in virtue of being the same kind of object, the dependence of the desirabilities of consequences on states is built into Jeffrey's framework. This means that his theory must dispense with the second of the restrictions required for Savage's.

The second modification that Jeffrey makes is more contentious. If he followed Savage in defining actions as arbitrary functions from partitions of events to consequences, the fact that in principle any proposition could serve as a consequence would imply an explosion in the size of the set of actions. But Jeffrey argues that many of the actions so defined would be inconsistent with the causal beliefs of the decision maker. Someone may think she has the option of making it true that if the traffic is light she will arrive on time for her appointment, and if it's heavy she will arrive late, but not believe that it is possible to make it true that if the traffic is light she arrives late, and if it's heavy she arrives on time. So, instead, Jeffrey conceives of actions as simply those propositions that can be made true at will, characterised for decision purposes by the probabilities and utilities (or desirabilities, as Jeffrey calls them) of the possible consequences that might be brought about by the action.

Two features of this treatment are noteworthy. Firstly, it is not required that the consequence of an action in each state be known in order that a decision be made. All that is required is that the agent have conditional probabilities, given the performance of the act, for the possible consequences of interest to her. This relaxes the first constraint on the applicability of Savage's theory.

Secondly, it is no longer required that the states of the world be probabilistically independent of the available actions. On the contrary, as Jeffrey sees it, actions matter precisely because they influence the probabilities of states (if you like, the consequences of acting *are* changed probabilities of states). This dispenses with the third constraint on the applicability of Savage's theory.

The fact that Jeffrey's theory imposes much weaker requirements on the framing of decision problems is my primary reason for preferring his framework to Savage's for developing a decision theory with a human face. There are other advantages too, such as the simplicity and flexibility of working with sets of propositions and the fact that the foundational representation theorems for his theory require much weaker assumptions about rational preference. But such flexibility does not come without cost. In particular, as we shall see, it opens up the question of exactly how acts should be evaluated, a matter of some controversy. So in the next part of the book I will develop a version of Bayesian decision theory that follows Jeffrey's in defining degrees of belief and desire on a common Boolean algebra of prospects (his propositions). But I will show how it is possible to extend the set of prospects in a way that allows for the re-introduction of Savage-style acts and a formulation of a state-dependent version of his theory. This richer theory is the one that I will defend as giving the best account of ideal rational agency.