2

Decision trees

2.1 INTRODUCTION

In the introduction of Chapter 1, I stated that the first task of a decision analyst was to try to obtain a coherent picture of her client's problem. In very simple problems, like Examples 1.1, 1.2 and 1.3, this may well be a trivial exercise but once problems start becoming more complex it can be very difficult to specify and quantify the relationships between decisions and uncertain events.

One well-known pictorial representation of a client's problem is the decision tree. Once drawn, the decision tree can be used to help calculate the expected pay-off associated with each sequence of decisions open to the client. It is then possible to identify the Bayes decision to a given problem.

The technique of drawing and 'folding back' a decision tree is most easily understood through an example. The following well-posed, if artificial, example is used to illustrate the method, and most of this chapter is devoted to its solution.

Example 2.1

The government has given an oil company the option to drill in either field A or field B but not both. The probabilities that oil is present in A and in B are 0.4 and 0.2 respectively and these two events are independent. A net profit of \$77 million is expected if oil is struck in A and a net profit of \$195 million is expected if oil is struck in B.

The decisions open to the company are:

(i) not to accept the option to drill in either area;
(ii) to accept the option and drill in either area A or area B immediately;
(iii) to pay for the investigation of one of the fields (but not both) and in the light of the information thus obtained, to choose between decisions (i) and (ii) above. The investigation will either advise drilling or not.

The investigative procedure, although costly, is not totally reliable. If oil is present in a field then the investigators will advise drilling with probability 0.8. If oil is not present, then the investigators will advise drilling with probability 0.4. The cost of accepting the option and drilling the chosen field is \$31 million while the cost of investigating a field is \$6 million.

Draw a decision tree and advise the company on the best decision rule under the expected monetary value algorithm. What is the expected pay-off to the company if they use this decision rule?

When given information in this form your first step should be to systematically tabulate all the information you have. In this example, let A and B denote the presence of oil in fields A and B respectively. Let $\bar{A}$ and $\bar{B}$ denote the complementary event to A and B. The events labelled a and b occur when the investigation advises drilling in respective fields A and B. The events labelled $\bar{a}$ and $\bar{b}$ will denote their complements.

Probabilities given in the problem Using the notation given above you have been given that

$$P(A)=0.4, \quad P(B)=0.2$$
$$P(a|A)=0.8, \quad P(b|B)=0.8, \quad P(a|\bar{A})=0.4, \quad P(b|\bar{B})=0.4$$

You are also told that A and B are independent, so clearly the result of an investigation of one field gives no information about the other field. So, for example, $P(a|A,B)=P(a|A)$. Formally stated, conditional on A or $\bar{A}$, a is independent of B, and conditional on B or $\bar{B}$, b is independent of A.

Thus we have, for example,

$$\begin{aligned} P(a\cap A\cap B) &= P(a|A\cap B)P(A\cap B)\\ &= P(a|A)P(A\cap B)\end{aligned}$$

by the conditional independence given above

$$= P(a|A)P(A)P(B)$$

by the independence of A and B. All probabilities relevant to your problem can be discovered using analogous formulae and are tabulated in Table 2.1.

Table 2.1 The joint probabilities of all events of interest in Example 2.1

Joint probabilities	$A\cap B$	$\bar{A}\cap B$	$A\cap\bar{B}$	$\bar{A}\cap\bar{B}$	Marginal probabilities
a	0.064	0.048	0.256	0.192	0.56
$\bar{a}$	0.016	0.072	0.064	0.288	0.44
b	0.064	0.096	0.128	0.192	0.48
$\bar{b}$	0.016	0.024	0.192	0.288	0.52
Marginal probabilities	0.08	0.12	0.32	0.48	

Table 2.2 Pay-offs (in \$ m) – profit less drilling cost

	Action		
Event	Drill A	Drill B	Don't drill
$A \cap B$	46	164	0
$\bar{A} \cap B$	−31	164	0
$A \cap \bar{B}$	46	−31	0
$\bar{A} \cap \bar{B}$	−31	−31	0

Pay-offs given in this problem In this problem there are three *terminal* decisions – that is, decisions which can be acted upon after choosing any preliminary investigation. These will be to drill A, to drill B or not to drill either field. The pay-off when making each decision given the states of the two fields is the profit less any drilling cost, and are summarized from the question in Table 2.2.

Finally, the cost of any investigation is \$6m. Having summarized your information you are now ready to draw your tree.

2.2 HOW TO DRAW A DECISION TREE

(i) Find a large piece of paper (the back of computer print-out is very useful for drawing decision trees). Work from the left to the right of the page. First identify the set of decisions D_1 your client needs to make before he can observe any of the outcomes of interest. In this example these decisions are:

d_1 – to investigate field A
d_2 – to investigate field B
d_3 – to drill A without investigation
d_4 – to drill B without investigation
d_5 – to not investigate and to drill neither field.

Whenever your client has a choice between a set of decisions such as D_1, draw a square box to represent D_1 and a fork emerging from that box for each possible decision. Each such square box is called a *decision node* and each such fork is called a *decision fork*.

Thus in this example, start drawing your decision tree as in Fig. 2.1.

(ii) Given that your client takes any one of these initial decisions he may then observe some event which was initially uncertain to him. In this example, after taking decision d_1 to investigate field A he will observe either a positive (a) or negative ($\bar{a}$) recommendation to drill. If some such uncertainty is resolved after taking a decision, draw a circle (called a *chance* node) at the

Decision Trees

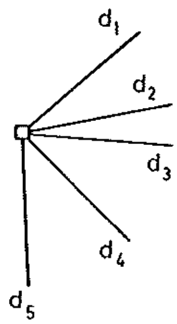

Figure 2.1

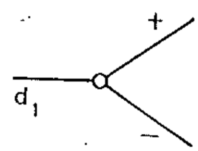

Figure 2.2 One of our initial chance nodes

right-hand end of your decision fork with forks (called *chance forks*) emerging to the right of it, each fork labelled by one of the possible outcomes. The first chance node to appear after decision d_1 of the example is drawn in Fig. 2.2.

(iii) You now draw on to the tree any further decision node that can represent a decision which needs to be made contingent only on decisions and outcomes that are already represented in the tree so far. For example, in the case above, given $\bar{a}$ your client must choose between drilling A, drilling B or not drilling. Draw another decision node with three forks out of it, labelled by these options. In general, include all logically feasible actions on the tree whether or not they, at first glance, seem sensible.

Similarly, you can add further chance nodes to the tree when the corresponding events have probabilities which only depend on decisions and outcomes you have already depicted and whose uncertainty is resolved after these previous decision and outcomes have occurred.

By adding nodes to the tree in this way you will eventually reach a point where no more nodes and decisions can be added using these rules. There may, however, be some residual uncertainty in the problem. In this example, after (d_1, a, drill A), there still remains to represent whether oil exists in A or not, and this will effect your expected pay-off. In this case this uncertainty can be represented by a single outcome node with two branches denoted by $A|a$ and $\bar{A}|a$. Adding sufficient outcome nodes to represent all your client's residual uncertainties, you will obtain the completed skeleton of your tree representing Example 2.1 as given in Fig. 2.3.

Now you have a diagram of your client's problem you should pause to ask yourself the question: 'Have I included all his possible options and included all aspects of uncertainty pertinent to choosing between those options?' Having drawn your tree you will often find that this is not the case. If you believe you have missed out some promising looking decisions or some large component of variation, add them to your model now by adding

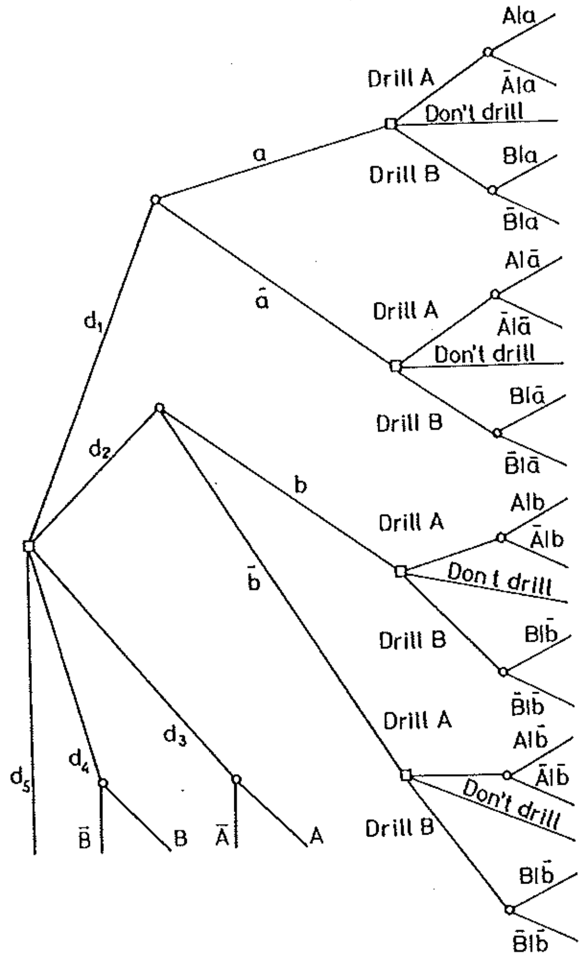

Figure 2.3 A tree skeleton

more decision nodes and forks and chance nodes and forks to your tree. Do this until you are happy that the tree is a good reflection of the problem in hand. Add to your summary of information all the necessary probabilities and pay-offs associated with these newly included options or components of variation.

(iv) Now your tree is drawn you need to fill your tree with the information provided. Firstly the pay-offs associated with each combination of outcome and decision need to be added to the right-hand tip of the *terminal* forks of your tree – that is, those forks with no node on their right-hand end. Care must be taken to ensure that you include all profits and costs incurred from taking decisions and observing outcomes that lead to that terminal fork. In our example, the pay-offs associated with the terminal nodes are just the corresponding relevant pay-offs obtained from Table 2.2 minus any cost of investigating an oilfield that might be incurred on the way to that terminal fork. Thus the pay-off associated with the sequence deciding to investigate A (d_1), getting a positive response (a), drilling A and finding oil (A) is \$46 m minus the \$6 m cost of investigation, i.e. \$40 m. Write the number 40 on the right-hand tip of the $A|a$ fork. If you proceed in this way for all the terminal forks of Fig. 2.3 you should obtain the 'terminal pay-offs' given in Fig. 2.4.

(v) The last components you need to add to your tree are the probabilities

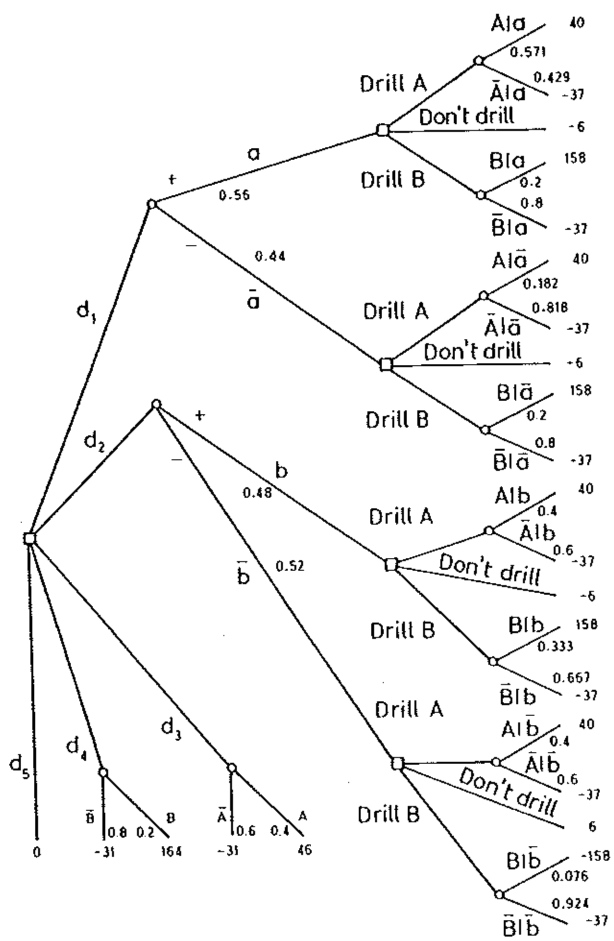

Figure 2.4 A tree with completed utilities and probabilities

relevant to the chance forks of the tree. In this example, suppose your client is considering decision d_1 of whether to investigate oilfield A. At the time of this decision your client is uncertain about whether he will observe a positive reading. (a) or not ($\bar{a}$). However, you have probabilistic information about the event a. From Table 2.1, $P(a)=0.56$ and $P(\bar{a})=0.44$. Write these probabilities over the relevant chance forks. Now let's suppose you observe a positive result and your client then decides to drill field A. This is depicted in Fig. 2.3 by the two topmost terminal chance forks. You need to know the probability of A (and hence $\bar{A}$) *given* these previous outcomes and decisions. In this case the only thing that influences this probability is a. $P(A|a)$ is not given directly but can be easily calculated, either from the original data using Bayes rule,

$$P(A|a)=\frac{P(a|A)P(A)}{P(a)}$$

or by using the probabilities given in Table 2.1 using the formula

$$P(A|a)=\frac{P(A\cap B\cap a)+P(A\cap \bar{B}\cap a)}{P(a)}$$

In either case you will calculate $P(A|a)=0.571$ and hence $P(\bar{A}|a)=1-P(A|a)=0.429$. These probabilities can now be added to the uppermost and next to uppermost terminal chance forks respectively.

Probabilities need to be added to each chance fork in this way. If you do this you should obtain the numbers given in Fig. 2.4.

(vi) Your tree now depicts the structure of the problem and contains all information relevant to finding a Bayes decision. Here is the way you identify your client's Bayes decision. It is called 'folding back the tree'.

Work from right to left. If there are any terminal chance forks then you write an expected pay-off over the corresponding chance node to the left of these forks. This expected pay-off is simply the product of terminal pay-off and associated probability summed over this set of forks. For example, over the chance node arising from d_1, a, Drill A, write the number 7.00 where 7.00 is calculated from

$$0.571\times 40+0.429\times(-37)=7.00$$

(vii) By doing this for all terminal chance forks you will have a set of decision forks with either a terminal pay-off at its right-hand end (as for d_5) or a chance node with an associated expected pay-off. Choose the decision fork from a decision node D' with the largest pay-off/expected pay-off, write this expected pay-off over D' and delete the other forks that emerge out of D'. For example, for the decision node after d_1, a, the pay-offs are 7 (Drill A),

suppose your client
ld A. At the time of
ll observe a positive
information about
= 0.44. Write these
ippose you observe
A. This is depicted
u need to know the
omes and decisions.
y is a. $P(A|a)$ is not
original data using

: formula

d hence $P(\bar{A}|a) =$
d to the uppermost
y.
this way. If you do

n and contains all
he way you identify
: tree'.
nce forks then you
node to the left of
of terminal pay-off
For example, over
number 7.00 where

ive a set of decision
nd (as for d_5) or a
e the decision fork
l pay-off, write this
emerge out of D'.
offs are 7 (Drill A),

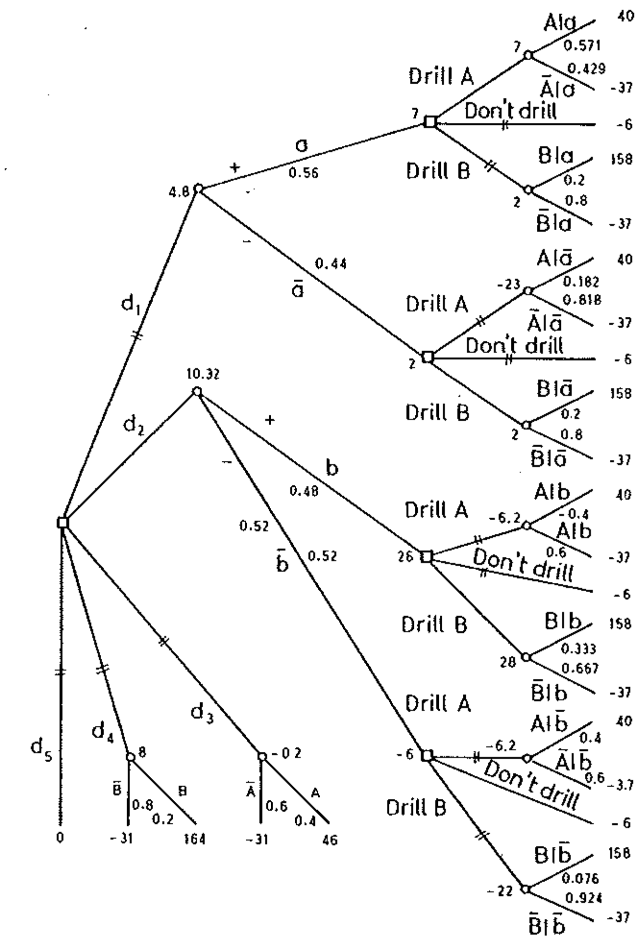

Figure 2.5 A 'folded back' decision tree

-6 (Don't drill) and 2 (Drill B); so delete the Don't drill and Drill B forks and write 7 over the decision node.

(viii) Repeat procedures (vi) and (vii) so that numbers are finally written over nodes connected by a fork to the original decision node. The tree has now been completed. Figure 2.5 gives the decision tree from Example 2.1.

(ix) Your client's optimal decision can now be read from the tree. Pick the decision in the initial decision node on the left with associated fork having the largest expected pay-off. In this example this is decision d_2. Provided your client follows the ensuing course of action, his expected pay-off associated with this decision d_2 is exactly the expected pay-off written over its right-hand node (\$10.32 m).

Of course it is not enough in this problem to just state that your client should investigate field B (decision d_2). Your client also needs to know how to act after making this decision. This can also be read from the tree. In this example, having taken decision d_2 you will discover the result of your investigation (b or $\bar{b}$). Given outcome b, your client should act on the decision labelled by the undeleted decision fork. In our tree this is the 'Drill B' fork. On the other hand if the investigation advises against drilling (outcome $\bar{b}$) your client should take the course of action 'Don't drill either field' labelling the undeleted fork arising from this outcome.

This optimal course of action, which specifies how your client should act contingent on the outcomes of various events which at the time of analysis are uncertain, is called a *Bayes decision rule.*

2.3 WHY THIS ALGORITHM FINDS THE DECISION RULE WHICH MAXIMIZES EXPECTED PAY-OFF

Here we interpret what the various forks and nodes represent in a technical sense and hence explain why the algorithm given in the last section actually identifies the optimal decision rule.

For any node n connected on the left to forks $f_1, \ldots, f_m$, the number above the right-hand end of f_i denotes the expected pay-off $G(f_i|\boldsymbol{\theta}, \mathbf{d})$ that is:

(a) conditional on all outcomes $\boldsymbol{\theta}$ occurring and decisions $\mathbf{d}$ made and represented by forks 'preceding' f_i, i.e. lying along a line connecting the far left-hand node at the base of the tree to f_i;

(b1) conditional on, in the case of n being a decision node, the decision labelled by f_i, $1 \leqslant i \leqslant m$;

(b2) conditional on, in the case of n being a chance node, the outcome θ_i occurring that is labelled by f_i, $1 \leqslant i \leqslant m$; and

(c) provided that your client subsequently takes the decisions available to him which will maximize his expected pay-off conditional on $(\boldsymbol{\theta}, \mathbf{d})$.

At a decision node therefore, our algorithm simply picks out the fork denoting the decision which promises the client his maximum expected pay-off provided that he subsequently acts wisely. This expected pay-off is placed over n. On the other hand, when n is a chance node our algorithm just calculates a conditional expected pay-off given subsequent wise actions using a well-known property of expectation. Thus the expected pay-off $\bar{R}(\boldsymbol{\theta}, \mathbf{d})$ given $\boldsymbol{\theta}, \mathbf{d}$ (the number to be placed over n) can be calculated by the formula

$$\bar{R}(\boldsymbol{\theta}, \mathbf{d}) = \sum_{i=1}^{m} \bar{R}(f_i|\boldsymbol{\theta}, \mathbf{d})P(\theta_i|\boldsymbol{\theta}, \mathbf{d})$$

where $P(\theta_i | \theta, \mathbf{d})$ is the probability of θ_i given the outcomes and decision leading to it.

Using the interpretation of the number calculated above, it is not difficult to prove inductively that our algorithm chooses at each decision fork, the decision which maximizes expected pay-off conditional on all preceding outcomes and decisions.

For a more detailed and formal proof that this algorithm maximizes expected pay-off, see Raiffa and Schlaifer (1961).

2.4 THE EXPECTED VALUE OF PERFECT INFORMATION

Before drawing a tree it is often a good idea to calculate the expected value of perfect information (EVPI). This is defined as your client's expected pay-off given he knows the state of nature minus his expected pay-off given he performs no experimentation. In our example you can see from Tables 2.1 and 2.2 that his expected pay-off given perfect information about the existence of oil, or otherwise, in the two fields, given that your client chooses a decision to maximize his expected pay-off, is:

$$0.08 \times 164 + 0.12 \times 164 + 0.32 \times 46 + 0.48 \times 0 = 47.5 \qquad \text{(in \$m)}$$

His best expected pay-off given that he does not experiment, i.e. investigate one of the oil fields, is \$8 m, the expected pay-off of drilling B immediately. Thus

$$\text{EVPI} = 47.5 - 8 = 39.5 \qquad \text{(in \$m)}$$

Any decision rule which performs an investigation that costs \$$C$ where $C > 39.5$ can be immediately discarded. This is because, even if the investigation provided perfect information about the two fields, the pay-off would still be less than \$8 m and so less preferable to drilling B immediately.

If you calculate the EVPI before drawing a decision tree you will often save time analysing decision rules which are clearly not going to be optimal (see Exercise 2.3). The EVPI gives you an upper bound to the amount your client should be prepared to consider paying for information.

2.5 PRACTICAL PROBLEMS ASSOCIATED WITH DECISION TREES AND THE EMV ALGORITHM

1. In practical problems it is very difficult to isolate *all* the viable courses of action open to your client. In fact if you were pedantic and tried to analyse every conceivable possible course of action you would never have the time to finish any analysis. In practice you are advised to analyse the decision rules which seem to you likely to be close to optimal schemes. Of course,

there is always the danger if you do this that you will not find the decision rule which would be optimal had you analysed a large class of decision rules. However, this will not matter too much if the expected pay-off from your chosen decision is not much less than the expected pay-off from the optimal decision associated with the larger problem.

2. In practice both the pay-offs associated with various courses of action and outcomes and the probabilities of outcomes are often very difficult to assess. A considerable amount of recent research has been directed into trying to solve these problems and will be discussed in much more detail in Chapter 4 and 5. Of course, if the values you assign to your client's pay-offs and probabilities are inappropriate then you may well be giving him bad advice. So great care is needed here.

3. We assume in this analysis that the EMV criterion of maximizing the company's expected pay-off was an appropriate one to use. Unfortunately this is often not the case. In the next chapter we develop ways of defining an optimal decision when it is inappropriate to use the EMV criterion.

EXERCISES

2.1 On his twentieth birthday a patient is brought into a hospital with an illness which is either Type I, with probability 0.4, or Type II, with probability 0.6. Independent of the type of illness, without treatment he will die on that day with probability 0.8 and otherwise survive and have normal life expectancy.

The surgeon has three possible courses of action open to him:

(i) not to treat the patient;
(ii) to give the patient drug L once;
(iii) to operate on the patient once.

He cannot both operate and administer the drug.

Both operating and administering the drug are dangerous to the patient. Independently of the type of illness, operating on the patient will kill him with probability 0.5 and the drug will kill him with probability 0.2.

If the patient survives the poisonous effects of the drug, it will either cure him or have no effect, each with probability 0.5, if he has Type I illness; and will have no effect if he has Type II illness. If the patient survives an operation, it will cure him with probability 0.8 if he has Type I illness and with probability 0.4 if he has Type II illness, otherwise having no effect. Survival of the patient will give him a life expectancy of 70 years in all cases.

Draw a decision tree to represent this problem. Calculate the surgeon's best strategy assuming that he wishes to maximize his patient's life expectancy.

2.2 A company called 'Prune', which currently markets personal computers, will need to choose between at most three options in three years' time:

a_1 – continue to market its current machine (called Prunejuice);
a_2 – market an improved version of Prunejuice instead (called Pruneplus);
a_3 – market a much more powerful machine instead (called Superprune)

Prune can choose between a_1, a_2, a_3 but cannot implement more than one of them.

A machine can be marketed only if it has been researched and developed (R&D) successfully. The event that an R&D programme could be successful for Pruneplus and the event that it would be successful for Superprune are considered independent with current respective probabilities 0.9 and 0.6. If neither of the new machines has been successfully researched and developed in the next three years then Prune would be forced to choose option a_1 – to market its current machine. Prune now needs to choose whether to:

(i) R&D neither machine (decision d_1) at a cost of \$0;
(ii) R&D Pruneplus only (decision d_2) at a cost of \$3 000 000;
(iii) R&D Superprune only (decision d_3) at a cost of \$5 000 000;
(iv) R&D both Pruneplus and Superprune (decision d_4) at a cost of \$8 000 000.

Given successful R&D, Prune expects to make \$2 000 000 net profit from Prunejuice, \$10 000 000 net profit from Pruneplus and \$18 000 000 net profit from Superprune. Draw a decision tree representing Prune's *current* decision problem, and advise Prune on its best plan of action given that they want to maximize their expected net profit less costs.

2.3 A customer insists you give him a guarantee that a piece of machinery will not be faulty for one year. As the supplier you have the option of overhauling your machinery before delivering it to the customer (action a_2) or not (action a_1). The pay-offs (in \$) you will receive by taking these actions when the machine is, or is not, faulty are given below.

State	Action a_1	a_2
Not faulty	1000	800
Faulty	0	700

The machinery will work if a certain plate of metal is flat enough. You have scanning devices which sound an alarm if a small region of this plate is bumpy. The probabilities that any one of these scanning devices sound an alarm given that the machine is, or is not faulty, are respectively 0.9 and 0.4. You may decide to scan with n scanning devices (d_n), $n = 0, 1, 2, 3, \ldots$, at the cost of \$$50n$ and overhaul the machine or not depending on the number of alarms that ring. These devices will give independent readings conditional on whether the machine is faulty or not. Prior to scanning you believe that the probability that your machine is faulty is 0.2. Use the expected value of perfect information to determine those decision rules which might be optimal. Draw a decision tree of these decision sequences to find the optimal number of scanning devices to use and how to use them to maximize your expected pay-off.