5

Information and Informational Decisions

In Part I of this book, covering *the economics of uncertainty*, the key topics addressed were (1) individuals' optimizing "terminal" choices with regard to the bearing of risk, and (2) the market equilibrium determined by the aggregate of such individual decisions. In turning here in Part II to *the economics of information*, we will similarly be considering issues on individual and on group decision making. On the individual level, the central question is: supposing you could get additional information before having to make a terminal decision, how ought you to decide whether and how much information to collect? On the group level, the main question is: what is the nature of the overall equilibrium that arises when some or all individuals undertake such "informational" actions?

After some introductory discussions, the present chapter first analyzes individuals' decisions as to the acquisition of information. We then consider the informational choices of groups, since the necessity to act as a collectivity has significant consequences for the nature and amount of information desired. Later chapters take up the implications for market equilibrium.

5.1 Information – Some Conceptual Distinctions

"Information" is a word with many meanings. Some of the distinctions to be brought out here will be useful in what follows.

Information as knowledge versus information as news

Information is sometimes taken to mean *knowledge* – an accumulated body of data or evidence about the world. From this point of view, information is a stock magnitude. But the word may also denote an increment to this stock of knowledge, in the form of a *message* or an item of *news*.

Information versus beliefs

Terms like knowledge, information, news, etc., are generally understood to refer to objective evidence about the world. *Belief* is the subjective correlate of knowledge. In this chapter we will be seeing how increments of objective knowledge (news or messages) lead rational individuals to revise their beliefs. Ultimately, however, decisions must be based upon subjective beliefs.

News and message – messages versus message service

The term "message" is generally taken to mean an intended communication from one to another person. "News" is somewhat more general and may refer to evidence or data arrived at by some process other than interpersonal communication, for example, by observing the weather. (This is receiving a message from nature, so to speak.) But we will be treating the words "news" and "message" as synonymous. A more essential distinction for our purposes, one that we will be insisting upon, is between a "message" and a "message service" (or between "news" and a "news service"). Rothschild's legendary carrier pigeon, which supposedly brought him early news of the outcome at Waterloo, was a message service; the report of Napoleon's defeat was the message. Since you can never know in advance what you will be learning, you can never purchase a *message* but only a message service – a set of possible alternative messages.[1]

Communication: intended versus inadvertent, inaccurate versus deceptive

Apart from intended messages, there may, of course, be unintentional communications of various types: (i) a report intended for one party may be overheard by another (or even intercepted so as never to arrive at its intended destination). More important for our purposes is: (ii) attempting to use your knowledge may inadvertently reveal it. If you purchase stock in Universal Petroleum on the basis of information that they have just discovered a new oil field, your action tends to drive up the price of Universal stock, thereby providing a signal to other traders.

As for accuracy, the content of a message may, of course, be more or less incorrect. Or, even if fully accurate as sent, it may be garbled in transmission before receipt. (It may even be *intentionally* inaccurate as sent, which need not imply a deceptive purpose; the inaccuracy may even be necessary for

[1] We sometimes use "signal" to mean a "message" and "information signal" to mean a "message service."

conveying the truth, as when the laws of visual perspective require that lines be made to converge if they are to be perceived as parallel.)[2]

Public versus private information

One of the crucial factors affecting the economic value of information is its *scarcity*. At one extreme, a particular datum may be possessed by only a single individual (private information); at the other extreme, it may be known to everyone (public information). *Publication* is the conversion of information from private to public status. All dissemination, even confidential communication to a single other person, involves some loss of privacy and thus of any value attaching thereto. (It would be possible to transmit purely private information only if a technology were to emerge whereby *forgetting* could be reliably effectuated. Then, for a price, I might give you some news while arranging to forget it myself!)

Public versus private *possession* of information must not be confused with quite another point, how narrow or widespread is the *relevance* of a bit of news. Whether or not there is oil under my land is of special private relevance to me, but discovery of a cheaper oil-extraction process may affect almost everyone.

First-order information versus second-order information

First-order information is about events – the outcome of the battle of Waterloo, or the result of tossing a die. In contrast, higher-order information relates to the message or information itself. There are many different aspects of an item of news about which we might like to become informed, quite apart from its content: for example, its source (which may be a clue to accuracy), or its privacy (how many people already know it).

We can distinguish an *overt secret* from a *covert secret*. Someone interested in marketing an item of news – or, to state things more carefully, in selling a message service that will emit a not-yet-known news item as its message – of course, needs to keep the actual message quite secret in advance of sale. But he will be broadcasting the higher-order information that he has such a secret. In contrast, someone engaged in espionage may be as urgently concerned to conceal the higher-order information (his possession of a secret) as the content of the secret itself.

One special case of higher-order information is the condition called "common knowledge." This is said to exist, say, when two persons both

[2] See also Section 12.1.1, where the purpose of sending a garbled message is deception.

Table 5.1. *Consequences of terminal choices*

		States (s)			
		1	2	...	S
	1	c_{11}	c_{12}	...	c_{1S}
	2	c_{21}	c_{22}	...	c_{2S}
Acts (x)	...	...	...	...	...
	X	c_{X1}	c_{X2}	...	c_{XS}
	Beliefs:	π_1	π_2	...	π_S

know a certain fact, each knows the other knows it, each knows that the other knows he knows it, and so forth (Aumann, 1976). We will see below that having "concordant beliefs" (agreed estimates of the probabilities of different states of the world) does not in general lead to the same economic consequences as having those same beliefs as "common knowledge."

5.2 Informational Decision Analysis

This section analyzes an individual's optimizing choice between the alternatives of: (1) taking immediate terminal action, versus (2) acquiring better information first, with the aim of improving the ultimate terminal decision to be made.

5.2.1 The Use of Evidence to Revise Beliefs

Table 5.1 is a generalized version of the simple picture of Table 1.1 in the opening chapter. For a set of available *terminal* actions $x = (1, \ldots, X)$ and a set of states of the world $s = (1, \ldots, S)$, the individual's choice of action and nature's selection of the state interact to determine the associated consequence c_{xs}. The bottom margin of the table shows the distribution of the individual's current probability beliefs, where π_s is the probability attached to the occurrence of state s, and, of course, $\sum_s \pi_s = 1$.

In taking terminal action, a person will choose whichever act x has the highest expected utility for him:

$$\underset{(x)}{\text{Max}}\, U(x) \equiv \sum_s \pi_s v(c_{xs}) \tag{5.2.1}$$

Here $v(c)$ is, as before, the elementary utility function.

Table 5.2. *Joint probability matrix* ($J = [j_{sm}]$)

				Messages (m)	M	Probabilities for states
	J	1	2	...		
	1	j_{11}	j_{12}	...	j_{1M}	π_1
States	2	j_{21}	j_{22}	...	j_{2M}	π_2
(s)	...	...	...	...	...	
	S	j_{S1}	j_{S2}	...	j_{SM}	π_S
Probabilities for messages		q_1	q_2	...	q_M	1.0

Receipt of any particular message m will generally lead to a revision of probability beliefs, and thus may possibly imply a different choice of best terminal action. We now ask how the individual's probability estimates should be revised in the light of new information, that is, how he should convert his *prior* probabilities into *posterior* probabilities.

In this belief revision process, five different probability measures may be involved:

π_s = the unconditional or marginal (prior) probability of state s
q_m = the unconditional probability of receiving message m
j_{sm} = the joint probability of state s and message m
$q_{m \cdot s}$ = the conditional probability (or "likelihood") of message m, given state s
$\pi_{s \cdot m}$ = the conditional (posterior) probability of state s, given message m

There are a number of ways of displaying the interaction among the various probability distributions involved. The clearest is to start with the *joint probability matrix J* pictured in Table 5.2. In the main body of the table, j_{sm} is the joint probability of the state being s and the message being m. (For example, the state might be "rain tomorrow" and the message "barometer is falling.")

The sum of all these joint probabilities, taken over all the messages and states, is, of course, unity: $\sum_{s,m} j_{sm} = 1$. For each given state s, summing over the messages m (i.e., summing the j_{sm} horizontally in each row of the J matrix) generates the corresponding prior "marginal" state probabilities π_s shown in the adjoined column at the right of Table 5.2. This corresponds to the probability identity:

$$\sum_m j_{sm} \equiv \pi_s$$

Table 5.3. *Likelihood matrix* ($L \equiv [l_{sm}] \equiv [q_{m \cdot s}]$)

		Messages (m)				
	L	1	2	...	M	
	1	$q_{1\cdot 1}$	$q_{1\cdot 1}$	...	$q_{M\cdot 1}$	1.0
	2	$q_{1\cdot 2}$	$q_{2\cdot 2}$	...	$q_{M\cdot 2}$	1.0
States(s)	...	...	...	...	...	...
	S	$q_{1\cdot S}$	$q_{2\cdot S}$	...	$q_{M\cdot S}$	*1.0*

Similarly, of course, summing over the states s (i.e., adding up the j_{sm} vertically in each column) generates the "marginal" message probabilities q_m, as shown in the row adjoined at the bottom of Table 5.2:

$$\sum_s j_{sm} \equiv q_m$$

The q_m are the prior probabilities that the individual attaches to receiving the different messages. Since the state probabilities and the message probabilities each make up a probability distribution – the grand sum taken either over the π_s or over the q_m – must again be 1.0, as indicated in the lower-right corner of the table.

Two other important matrices are readily derived from the underlying joint probability matrix J. We will call them the *likelihood matrix* L and the *potential posterior matrix* Π.

The likelihood matrix $L = [l_{sm}]$ (Table 5.3) shows the *conditional* probability of any message given any state, which will be denoted $q_{m\cdot s}$. Thus:

$$l_{sm} \equiv q_{m\cdot s} \equiv \frac{j_{sm}}{\pi_s} \tag{5.2.2}$$

Numerically, the elements of the L matrix are obtained from the J matrix by dividing the j_{sm} in each row through by the adjoined π_s. For any row of the L matrix, these conditional probabilities must, of course, sum to unity, since:

$$\sum_m q_{m\cdot s} \equiv \frac{1}{\pi_s} \sum_m j_{sm} \equiv \frac{1}{\pi_s} \pi_s \equiv 1$$

(But note that the *column* sums in the L matrix do not sum to unity, except by accident, and in fact these column sums have no meaning so far as our analysis is concerned.)

Table 5.4. *Potential posterior matrix* ($\Pi \equiv [\pi_{s \cdot m}]$)

		Messages (m)			
	Π	1	2	...	M
	1	$\pi_{1\cdot 1}$	$\pi_{1\cdot 2}$	...	$\pi_{1\cdot M}$
	2	$\pi_{1\cdot 2}$	$\pi_{2\cdot 2}$	...	$\pi_{2\cdot M}$
States (s)	...	...	...	...	...
	S	$\pi_{S\cdot 1}$	$\pi_{S\cdot 2}$	...	$\pi_{S\cdot M}$
		1.0	1.0	...	1.0

The potential posterior matrix Π (Table 5.4) shows the conditional probability of each state given any message m, which is denoted $\pi_{s\cdot m}$ and defined in:

$$\pi_{s\cdot m} \equiv \frac{j_{sm}}{q_m} \tag{5.2.3}$$

The elements of the Π matrix are obtained numerically from the underlying J matrix by dividing all the j_{sm} in each column through by the adjoined q_m below. The *column* sums must then all be unity, since:

$$\sum_s \pi_{s\cdot m} \equiv \frac{1}{q_m} \sum_s j_{sm} \equiv \frac{1}{q_m} q_m \equiv 1$$

(Here the *row* sums do not in general equal unity, and in fact have no relevant meaning for our purposes.) Why we term this the potential posterior matrix should be evident. Each separate column shows the "posterior" probability distribution for states of the world that a person with a given J matrix should logically adopt, after having received the particular message m. The entire Π matrix therefore gives us an *ex ante* picture of *all* the alternative posterior distributions that could come about, depending upon which of the possible messages is received.

Looking at this in a somewhat different way, the prior probability π_s of state s is an average, weighted by the message probabilities q_m, of the posterior probabilities $\pi_{s\cdot m}$ for state s:

$$\pi_s = j_{s1} + \cdots + j_{sM} = q_1 \pi_{s\cdot 1} + \cdots + q_M \pi_{s\cdot M}$$

Or, expressing the relation between the prior state-probability vector $\pi = (\pi_1, \ldots, \pi_s)$ and the message-probability vector $q = (q_1, \ldots, q_M)$ in matrix notation:

$$\pi = \Pi q \tag{5.2.4}$$

And analogously, the message probability q_m is an average of the prior state probabilities weighted by the likelihoods:

$$q_m = j_{1m} + \cdots + j_{sm} = \pi_1 q_{m\cdot 1} + \cdots + \pi_S q_{m\cdot S}$$

So that, in matrix notation:

$$q = L'\pi$$

Example 5.1: There may be oil under your land. The alternative terminal actions are to undertake a major investment to develop the field or not to do so, and this decision will, of course, depend upon your estimate of the chance of oil really being there. The underlying "states of the world" are three possible geological configurations: state 1 is very favorable, with 90% chance that oil is there; state 2 is much less favorable, with 30% chance; state 3 is hopeless, with 0% chance.

In order to improve your information before taking terminal action, you have decided to drill a test well. The two possible sample outcomes or messages are that the test well is either "wet" or "dry." On the assumption that the test well is a random sample, then, if the true state is really state 1, there is a 90% chance of the message "wet"; if it is state 2, there is a 30% chance; and if it is state 3, no chance. So the given data specify the likelihood matrix $L \equiv [q_{m\cdot s}]$ shown below.

Suppose that, in addition, you initially attach probabilities $(\pi_1, \pi_2, \pi_3) = (0.1, 0.5, 0.4)$ to the three states. Multiplying each likelihood $q_{m.s}$ in the L matrix by the prior probability π_s yields the joint probability matrix $J \equiv [j_{sm}]$. The column sums are then the message probabilities $q_1 = 0.24$ and $q_2 = 0.76$. Using these you can easily compute your potential posterior matrix $\Pi = [\pi_{s\cdot m}]$.

		$L \equiv [q_{m\cdot s}]$ Messages (m)				$J \equiv [j_{sm}]$ Messages (m)				$\Pi \equiv [\pi_{s\cdot m}]$ Messages (m)	
		Wet	Dry			Wet	Dry			Wet	Dry
	L	1	2		J	1	2	π_s	Π	1	2
	1	0.9	0.1	1.0	1	0.09	0.01	0.1	1	0.375	0.013
States	2	0.3	0.7	1.0	2	0.15	0.35	0.5	2	0.625	0.461
(s)	3	0	1.0	1.0	3	0	0.40	0.4	3	0	0.526
					q_m	0.24	0.76	1.0		1.0	1.0

Notice that the message "wet" shifts your prior distribution $(\pi_1, \pi_2, \pi_3) = (0.1, 0.5, 0.4)$ to the much more favorable posterior distribution $(\pi_{1 \cdot 1}, \pi_{2 \cdot 1}, \pi_{3 \cdot 1}) = (0.375, 0.625, 0)$. The message "dry" leads, of course, to a much less hopeful posterior distribution $(\pi_{1 \cdot 2}, \pi_{2 \cdot 2}, \pi_{3 \cdot 2}) = (0.013, 0.461, 0.526)$.[3] □

So far as pure logic is concerned, the relevant data might equally well be presented or summarized in three different ways: (1) by the joint probability matrix J; (2) by the prior probability distribution for *states* (the right-hand margin of the J matrix) together with the likelihood matrix L; and, (3) by the prior probability distribution for *messages* (the bottom margin of the J matrix) together with the potential posterior matrix Π. But from the operational point of view, a decision maker will usually find it most convenient to use method 2. As in the example above, he is likely to think in terms of having a prior probability distribution for the underlying states together with a likelihood matrix L describing the message service being employed.

Prior beliefs are always *subjective*, while the likelihood matrix summarizing the possible message outcomes may often (though not necessarily) be *objective*. Suppose the message service represents the outcome of two tosses of a coin. For the state of the world that the coin is fair, with unbiased sampling the likelihood of the message "two heads" is objectively calculable from the laws of probability as 1/4. But someone who is not sure that the sampling is unbiased might subjectively assign a somewhat different likelihood to the message "two heads."

This process of revision of probabilities is called *Bayesian*, after Bayes' Theorem. The derivation is simple. First, the joint probability j_{sm} can be expressed in two ways in terms of the conditional probabilities defined in (5.2.2) and (5.2.3):

$$\pi_s q_{m \cdot s} \equiv j_{sm} \equiv q_m \pi_{s \cdot m}$$

[3] Random or "unbiased" sampling was assumed here, e.g., in state 1 (defined in terms of a 90% chance of oil being there) the likelihood matrix L indicates a 90% probability $q_{1.1}$ of receiving the message "wet." The underlying method is, however, sufficiently general to allow even for biased sampling. Thus, if state 1 is the true state of the world, conceivably the well (perhaps because it will not be drilled to full depth) could still have only an 80% chance of showing "wet." Whatever the likelihoods are, they can be displayed in the L matrix (and equivalently reflected in the J and Π matrices as well).

(Note how these three formulations relate to the information contained in the L, J, and Π matrices, respectively.) Solving for the posterior probability $\pi_{s \cdot m}$:

Bayes' Theorem (I)

$$\pi_{s \cdot m} \equiv \frac{j_{sm}}{q_m} \equiv \pi_s \frac{q_{m \cdot s}}{q_m} \tag{5.2.5}$$

In words: The posterior probability that an individual should attach to state s, after receiving message m, is equal to the prior probability π_s multiplied by the likelihood $q_{m.s}$ of message m, and then divided by a normalizing factor, which is the overall probability q_m of receiving message m. Since the latter is the "marginal" probability of m, we can also write Bayes' Theorem in the alternative forms:

Bayes' Theorem (II)

$$\pi_{s \cdot m} \equiv \pi_s \frac{q_{m \cdot s}}{\sum_{s'} j_{s'm}} \equiv \pi_s \frac{q_{m \cdot s}}{\sum_{s'} \pi_{s'} q_{m \cdot s'}} \tag{5.2.5}$$

The Bayesian belief-revision process is illustrated in Figure 5.1. For diagrammatic convenience, the pictorial representation assumes a continuous rather than a discrete state-defining variable, running from a lower limit $s = 0$ to an upper limit $s = S$. The "prior probability" curve in the diagrams is the given initial probability density function $\pi(s)$, where, of course, $\int_0^S \pi(s)ds \equiv 1$. By assumption here, this density shows some humping toward the middle; i.e., prior beliefs are such that middling values of s are regarded as more likely than extreme ones.

There are two alternative messages: $m = 1$ has much greater likelihood if the true state of the world is toward the high end of the range (upper diagram), while $m = 2$ has greater likelihood if s is small (lower diagram). Although the prior probability distributions are the same in the upper and lower panels, the differing "likelihood function" curves lead to differing "posterior probability" density functions. In each panel, the posterior distribution is a kind of compromise or average of the other two curves. More specifically, for any s, the height $\pi_{s \cdot m}$ along the posterior probability density curve is the product of the height along the prior density curve (π_s) times the height along the likelihood function ($q_{m.s}$) – adjusted by a normalization or re-scaling factor to make the new integrated probability equal to unity. (This normalization corresponds to dividing through by the message probability q_m, as indicated in Equation (5.2.5).)

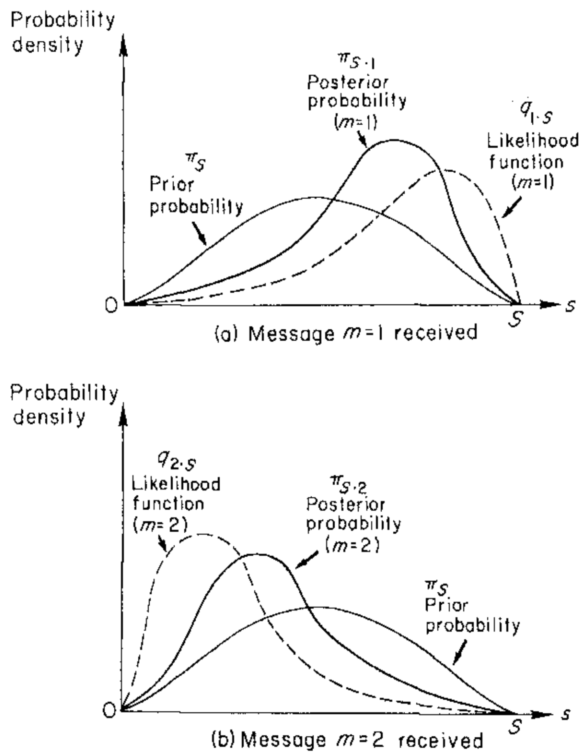

Figure 5.1. Bayesian revision of probability beliefs.

Three useful propositions are implicit in the Bayesian belief-revision process:

(1) Recall our discussion in the initial chapter about the individual's *confidence* in his beliefs, and how this would affect his decision whether or not to acquire more information. Confidence is indicated in Figure 5.1 by the "tightness" of the prior probability distribution – the degree to which a person approaches assigning 100% probability to some single possible value of s. The corresponding proposition is: other things equal, the higher the prior confidence, the more the posterior distribution will resemble the prior distribution for any given message or weight of evidence. This is intuitively clear if we go to the limit. An individual with "absolute" prior confidence, who attaches 100% probability to some particular value $s = \hat{s}$, can learn nothing from new evidence. From Equation (5.2.4), if the prior probabilities π_s all

equal zero for any $s \neq \hat{s}$, the corresponding posterior probabilities $\pi_{s \cdot m} = 0$ also.

(2) Other things equal, the greater the mass of new evidence the more the posterior distribution will resemble the likelihood function rather than the prior probabilities. A large sample size, for example, will be reflected in a very "tight" likelihood curve in the diagrams. Suppose that, in tossing a coin, P is the unknown true probability of obtaining heads while p is the sample proportion of heads observed in n tosses. Then $p = 0.5$ in a sample of two tosses would be reflected in a "loose" likelihood curve with only a mild peak around $P = 0.5$ – whereas $p = 0.5$ in a sample of 100 tosses would produce a very tight likelihood curve with a sharp peak around $P = 0.5$. As before, the tighter the curve, the more pull it has upon the shape of the posterior probability distribution.

(3) Other things equal, the more "surprising" the evidence, the bigger the impact upon the posterior probabilities. Intuitively, this is obvious: only when a message is surprising does it call for any drastic change in our beliefs. In terms of Equation (5.2.5), a "surprising" message would be one with low message probability q_m. Other things equal, the smaller the q_m in the denominator on the right-hand side, the bigger the multiplier causing the posterior probability $\pi_{s \cdot m}$ to diverge from the prior π_s.

Exercises and Excursions 5.2.1

1 The Game Show

A contestant on a television game show may choose any of three curtained booths, one of which contains a valuable prize. Lacking any prior information, she arbitrarily selects one of the booths, say, no. 1. But before the curtain is drawn revealing whether she wins or loses, the Master of Ceremonies says: "Wait, I'll give you a chance to change your mind." He then draws the curtain on one of the *other* booths, say, no. 2, which is revealed to be empty. The M.C. then asks if the contestant cares to change her choice. Should she do so?

2 Joint (J), Likelihood (L), and Potential Posterior (Π) Matrices

(A) There are two states s_1 and s_2 and two messages m_1 and m_2. The prior probability distribution over states is $\pi = (0.7, 0.3)$. The posterior probabilities $\pi_{s \cdot m}$ include $\pi_{1 \cdot 1} = 0.9$ and $\pi_{2 \cdot 2} = 0.8$. Calculate the Π, L, and J matrices.

(B) (i) If you have the J matrix, what additional data (if any) are needed to construct each of the other two matrices?
(ii) Same question, if you have the L matrix.
(iii) Same question, with the Π matrix.

(C) For the prior probability distribution π above, and still assuming only two possible messages, show (if it is possible to do so):
(i) An L matrix representing a *completely conclusive* message service. (That is, a matrix leaving no posterior uncertainty whatsoever.) If it exists, is it *unique*, or are there other such L matrices? Also, does any such matrix depend at all upon π?
(ii) An L matrix that is *completely uninformative*. Answer the same questions.
(iii) An L matrix that (if one message is received) will conclusively establish that one of the states will occur but (if the other message is received) will be completely uninformative. Same questions.

5.2.2 Revision of Optimal Action and the Worth of Information

If immediate terminal action is to be taken, the individual will choose whichever act has highest expected utility, as indicated in Equation (5.2.1). In condensed notation:

$$\operatorname*{Max}_{(x)} U(x; \pi) \equiv \sum_s \pi_s v(c_{xs}) \tag{5.2.6}$$

Denote as x_0 the optimal immediate terminal action, which, of course, must be calculated in terms of the prior probabilities π_s. What we are concerned with here is the value of an *informational action*, that is, the expected utility gain from using an information service.

After a particular message m has been received from such a service, the decision maker would use (5.2.6) once again, employing now the *posterior* probabilities $\pi_{s\cdot m}$. This recalculation could well lead to a choice of a different optimal terminal action x_m. Then ω_m, the value in utility units of the message m, can be defined as:

$$\omega_m \equiv U(x_m; \pi_{\cdot m}) - U(x_0; \pi_{\cdot m})$$

This is the expected gain from the revision of optimal action, calculated in terms of the individual's *revised* probabilities $\pi_{\cdot m} \equiv (\pi_{1\cdot m}, \ldots, \pi_{S\cdot m})$. Evidently, the expected utility gain must be non-negative, else x_m could not have been the optimal act using the posterior probabilities. (But message m

could lead to *no change* of best action despite the revision of probability beliefs, in which case $\omega_m = 0$ – such a message has zero value.)

However, one cannot purchase a given message but only a *message service*. So it is not the ω_m associated with some particular message that is relevant but rather the expectation of the utility gain from all possible messages weighted by their respective message probabilities q_m. More specifically, for a message service μ characterized by a particular likelihood matrix L and prior beliefs, π, the expected value of the information (the worth of the message service) is:

$$\Omega(\mu) = \mathop{\mathrm{E}}_{m}[\omega_m] = \sum_m q_m[U(x_m; \pi_{\cdot m}) - U(x_0; \pi_{\cdot m})]$$

Since each ω_m is non-negative, we know that a message service can never lower the agent's expected utility (before allowing for the *cost* of the service).[4]

Let c^*_{sm} denote the income in state s associated with the best action x_m after receiving message m, and c^*_{s0} the corresponding income for the best uninformed action x_0 (that is, the best action in terms of the prior beliefs):

$$\begin{aligned}\Omega(\mu) &= \sum_m q_m \sum_s \pi_{s \cdot m} v\left(c^*_{sm}\right) - \sum_m \sum_s \pi_{s \cdot m} q_m v\left(c^*_{s0}\right) \\ &= \sum_m \sum_s \pi_{s \cdot m} q_m v\left(c^*_{sm}\right) - \sum_s \pi_s v\left(c^*_{s0}\right) \qquad (5.2.7)\end{aligned}$$

Thus the value of the message service is just the difference between expected utility with and without the service.

For a simplified model with only two states of the world ($s = 1, 2$), Figure 5.2 illustrates a situation with three available terminal actions ($x = 1, 2, 3$). In this three-dimensional diagram, utility is measured vertically, while the probabilities of the two states are scaled along the two horizontal axes. Each possible assignment of probabilities to states is represented by a point along AB, a line in the base plane whose equation is simply $\pi_1 + \pi_2 = 1$.

Suppose state 1 occurs. Then the cardinal preference-scaling values (the elementary utilities) associated with the consequences c_{x1} attaching to the different actions x are indicated by the intercepts labeled $v(c_{11})$, $v(c_{21})$, and $v(c_{31})$ lying vertically above point B in the diagram. Similarly, $v(c_{12})$, $v(c_{22})$, and $v(c_{32})$, the elementary utilities of outcomes in state 2, are the

[4] This is true only in single-person settings. As we shall see in Section 5.3, in multi-person settings, one person's decision to acquire a message service may change another person's behavior. This can lead to a smaller expected utility to the person who acquires the message services.

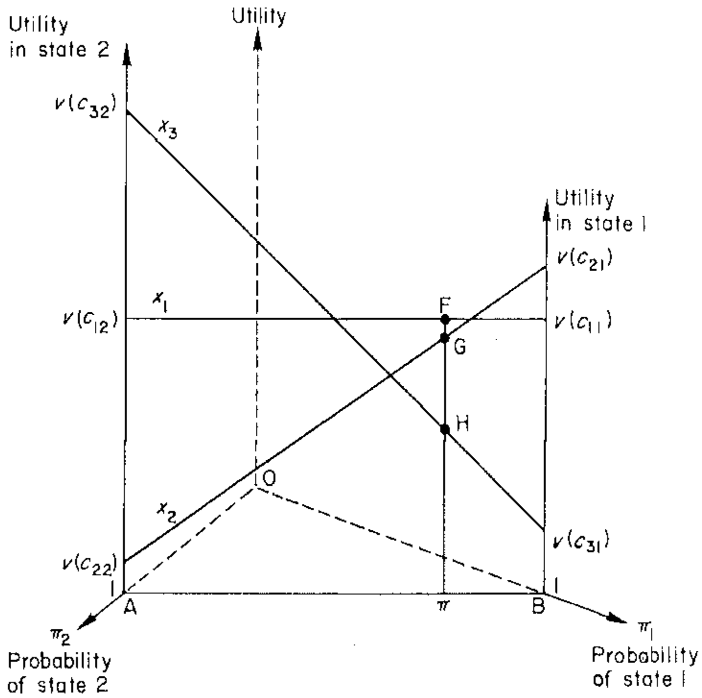

Figure 5.2. Best action in a three-action, two-state world.

corresponding intercepts above point A. (Note that x_1 is a certainty action, since it yields the same elementary utility in either state.) The expected utility $U(x_i, \pi) = \pi_1 v(c_{i1}) + \pi_2 v(c_{i2})$ of any action x, given any probability vector π, is shown by the vertical distance from the point $\pi = (\pi_1, \pi_2)$ along AB to the line joining $v(c_{x1})$ and $v(c_{x2})$ for that action. In the diagram, if π is the prior probability vector then the best immediate *terminal action* is the certainty action $x = 1$ whose expected utility is indicated by the height of point F above the base plane.

In Figure 5.2, everything of interest takes place in the vertical plane overlying the line AB. So we can simplify matters by shifting to a two-dimensional diagram, as in Figure 5.3, where the line AB becomes the horizontal axis along which π_1 varies from zero to unity. In this simplified diagram, we see once more the expected utilities of actions $x = 1, 2, 3$ as functions of the prior probability vector $\pi = (\pi_1, \pi_2)$, leading to the choice of $x = 1$ as the best terminal action in terms of the prior probabilities.

We now want to picture the effects of receiving information. Suppose an information service μ can generate two possible messages $m = 1, 2$. Either

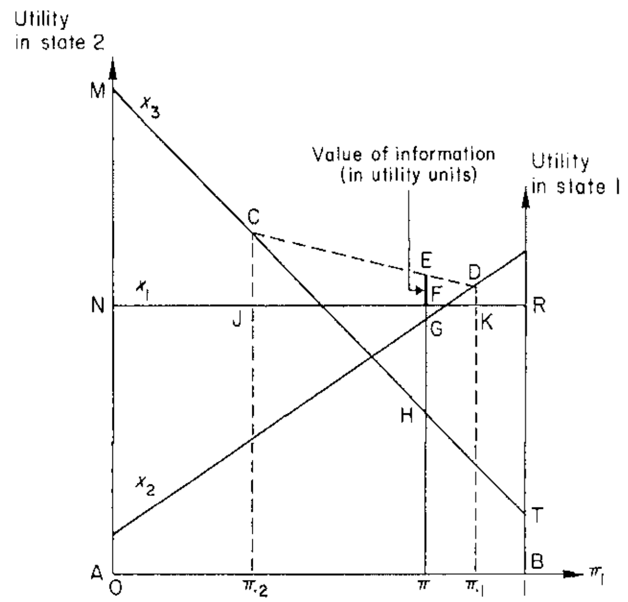

Figure 5.3. The value of information.

message will lead to a revised (posterior) probability vector $\pi_{\cdot m} \equiv (\pi_{1 \cdot m}, \pi_{2 \cdot m})$. If message $m = 2$ is received, by construction the revised choice of action in terms of the posterior probability vector $\pi_{\cdot 2} \equiv (\pi_{1.2}, \pi_{2.2})$ is $x = 3$, with expected utility indicated by the height of point C. The *ex post* expected-utility gain over $x = 1, \omega_1$, is therefore equal to the vertical distance CJ (the probability-weighted average of the *gain* MN if state 2 occurs and the *loss* RT if state 1 occurs). If, on the other hand, the message received is $m = 1$, the best action in the diagram is $x = 2$ (expected utility indicated by point D), so $\omega_2 = \text{DK}$. Then $\Omega(\mu)$, the value of the information service in terms of the gain in expected utility, is represented by the vertical distance EF above the point π along the line AB.

Example 5.2: In an oil-drilling situation, suppose there are just two possible states: "wet" (with prior probability $\pi_1 = 0.24$) and "dry" (with prior probability $\pi_2 = 0.76$).[5] If you take action $x = 1$ (drill the well) and it is wet, you gain \$1,000,000. If it is dry, you lose \$400,000. Action $x = 2$

[5] These numbers are consistent with those in the previous example. Although in that example there were three states, associated with three different possible geological formations, the implied overall probabilities worked out to 0.24 for wet and 0.76 for dry, as shown in the bottom row adjoined to the *J* matrix.

(not drilling) involves a \$50,000 cost in relocating your rig. Suppose your utility function is simply linear in income (you are risk neutral), so we can write $v(c) = c$. A message service, taking the form of a geological analysis in advance of drilling, is characterized by the following likelihood matrix L. How much should you be willing to pay for it?

		Message		
	$L = [q_{m\cdot s}]$	Wet	Dry	
State	Wet	0.6	0.4	1.0
	Dry	0.2	0.8	1.0

Answer: In terms of your prior probabilities, action $x = 1$ involves an expected gain of 0.24 (\$1,000,000)—0.76(\$400,000) = −\$64,000, whereas action $x = 2$ leads to a loss of only \$50,000. So the optimal prior action x_0 is $x = 2$. As for the value of the message service, straightforward computations lead to the Potential Posterior Matrix shown below.

		Message	
	$\Pi = [\pi_{s\cdot m}]$	Wet	Dry
State	Wet	0.486	0.136
	Dry	0.514	0.864
		1.0	1.0

Using the posterior probabilities, if the message is "dry", the best action remains $x = 2$ (not drilling). But if the message is "wet," the expected gain from drilling (action $x = 1$) becomes 0.486(\$1,000,000) − 0.514(\$400,000) = \$280,400. So the expected value of the information is 0.296 (\$280,400) + 0.704 (−\$50,000) − (−\$50,000) = \$97,798. This is the value of the message service, where 0.296 and 0.704 are the message probabilities q_1 and q_2. □

We have already seen, in discussing Bayes' Theorem, that higher prior confidence implies smaller revision of beliefs from given messages. It follows that higher confidence also implies lower value of information. With more confident prior beliefs, in Figure 5.3 the posterior probability vectors $\pi_{\cdot 1}$ and $\pi_{\cdot 2}$ would both lie closer to the original π. It is evident that the effect (if any) of greater prior confidence can only be to shrink the distance EF that represents the value of acquiring more evidence.

So far, we have discussed the value of information in utility units. The natural next step is to calculate what a message service is worth in income (corn) units – i.e., the maximum fee ξ that someone with utility function $v(c)$ would be willing to pay for the information.

As follows directly from Equation (5.2.7), the fee is determined in:

$$\sum_{m}\sum_{s}\pi_{s\cdot m}q_m v(c^*_{sm}-\xi)=\sum_{s}\pi_s v(c^*_{s0}) \tag{5.2.8}$$

That is, the maximum fee ξ that a person would pay, in advance of receiving the message, is such as to make the expected utility of the best informed action exactly equal to the expected utility of the best uninformed action.

Suppose that $x_m = x_0$ for all messages m. Then $c^*_{sm} = c^*_{s0}$ for all messages m, and $\xi = 0$ is the solution to (5.2.8). Thus, as one would expect, if the optimal action remains the same regardless of the message received, then the information service is worthless.

Example 5.3: In a two-state world, suppose the contingent-claim prices for corn in states 1 and 2 are numerically equal to the prior state probabilities: $p_1 = \pi_1$ and $p_2 = \pi_2$, where $\pi_1 + \pi_2 = 1$. (Thus, the prices are "fair.") Specifically, suppose the states are equally probable. Consider an individual with utility function $v(c) = \sqrt{c}$ and endowment $(\bar{c}_1, \bar{c}_2) = (50, 150)$.

Before acquiring any additional information, there are an infinite number of possible actions the individual might take, representing the possible amounts of his endowed c_1 he might trade for c_2, or vice versa. But, since the prices are fair, in accordance with Fundamental Theorem of Risk Bearing we know that the optimal action will be to trade to the certainty position $(c^*_{10}, c^*_{20}) = (100, 100)$. (The 0 subscript here signifies, as before, the best *uninformed* action.) In terms of a picture like Figure 5.3, the "null action" $\bar{x}$ – remaining at the endowment position – would be represented by a line with negative slope, the left intercept being at $\sqrt{150}$ and the right intercept at $\sqrt{50}$. The optimal uninformed action x^*_0 would be a horizontal line at $v(100) = \sqrt{100} = 10$. The expected utility of the endowment position is $(\sqrt{150} + \sqrt{50})/2$ or about 9.66, while the expected utility of x^*_0 is, of course, 10. So 0.34 is the utility gain, over the endowment position, of the best action under uncertainty.

Now suppose a message service μ, whose output will be *conclusive* as to which state is going to obtain, becomes available in time for the individual to engage in state-claim trading at the same fair prices. To determine the value of μ in utility units we first have to find, for each possible message m, the individual's best informed action. If message 1 is received, obviously the individual will convert all of his endowed wealth into 200 units of c_1, and similarly into 200 units of c_2 if message 2 is received. Thus his expected utility will be $\sqrt{200}$ or 14.14 approximately.

To determine the maximum fee ξ he would be willing to pay, we can use Equation (5.2.8), which reduces here to the simple form:

$$v(200 - \xi) = v(100)$$

The solution is $\xi = 100$. Thus, the individual here would be willing to pay up to half of his endowed wealth for a conclusive message, arriving in time for market trading, telling him which state is going to obtain. □

Exercises and Excursions 5.2.2

1 Value of Information in a Simple Betting Problem

You have an opportunity to gamble on the toss of a coin. If your choice is correct, you win \$30, but if it is wrong, you lose \$50. Initially, you think it equally likely that the coin is two-headed, two-tailed, or fair. If you are risk neutral, so that your utility function can be written $v(c) = c$, how much should you be willing to pay to observe a sample of size 1?

2 Value of Information in a Two-Action Problem with Linear Costs

You are the receiving officer of a company that has received a large shipment of ordered goods. You must decide whether to accept (action $x = A$) or reject (action $x = R$) the shipment. Which you will want to do depends upon the unknown proportion defective P in the shipment (population). Your loss function is:

$$L(R,P) = \begin{cases} 0, & \text{for} \quad P \geq 0.04 \\ 100(0.04 - P), & \text{for} \quad P < 0.04 \end{cases}$$

$$L(A,P) = \begin{cases} 0, & \text{for} \quad P \leq 0.04 \\ 200(P - 0.04), & \text{for} \quad P > 0.04 \end{cases}$$

Your prior probability distribution is defined over four discrete values of P (states of the world):

Fraction defective (P)	Prior probability (π_s)
0.02	0.7
0.04	0.1
0.06	0.1
0.08	0.1
	1.0

(A) What is your best *prior* decision, in the absence of sample information? Assume that you are risk neutral, with utility function $v(c) = c$.

(B) How much should you be willing to pay for a sample of size 1?

3 Value of Information with Logarithmic Utility

Two individuals with endowed wealths W_1 and W_2 have the same utility function $v = \ln(c)$, the same probability beliefs π_s, and face the same state-claim prices P_s.

(A) Individual 1 will receive no additional information, but individual 2 will be receiving conclusive information revealing the true state before trading takes place. Show that the utility difference between them is given by:

$$U_2 - U_1 = \ln W_2 - \ln W_1 - \sum_s \pi_s \ln \pi_s$$

[HINT: Make use of a simple relationship between W and c_s that holds for the logarithmic utility function.]

(B) Using this result, show that an uninformed individual with this utility function would be willing to give up a fraction K^* of his wealth to receive conclusive information before trading takes place, where:

$$K^* = 1 - (\pi_1)^{\pi_1} (\pi_2)^{\pi_2} \cdots (\pi_S)^{\pi_S}$$

(C) Show that he will pay the most when he initially assigns equal probabilities to all states.

4 Value of Less-than-Conclusive Information

Under the conditions of Example 5.3, suppose the message service μ does not provide a fully conclusive message as to which state is going to obtain. Instead, suppose message m_1 is such as to lead to the posterior distribution $(\pi_{1.1}, \pi_{2.1}) = (0.75, 0.25)$ while message m_2 leads to $(\pi_{1.2}, \pi_{2.2}) = (0.25, 0.75)$.

(A) Compute the likelihood matrix L associated with this message service.

(B) Show that the individual with the utility function in the text example, $v(c) = \sqrt{c}$, would be willing to pay a fee ξ equal to 20 units of corn for this information, provided as before that the message arrives in time for market trading.