5

Modeling Information

The information content perspective stresses the idea accounting is a source of information, one we hasten to add that uses the language and algebra of valuation to convey its information. To this point we have dealt with the initial portion of this recurring phrase, the language and algebra of valuation and its centrality in the "accounting model." It is now time to tackle the term information. This turns out to take some care, as the term information has taken on near colloquial status, yet being even moderately serious about its nature requires considerable structure, structure that describes what it is we are uncertain about and how the information speaks to that uncertainty.

The information content perspective stresses accounting's role as a source of information. To provide information means it is possible to become better informed, to learn something we did not yet know. In turn, being able to become better informed means we were initially less informed, that uncertainty was present. The implication should be understood: to treat accounting as a source of information demands we become facile with the notions of uncertainty and information.[1]

We begin with the notion of uncertainty. This entails how we represent resource allocation exercises when uncertainty is present. From here we move on to the question of how we represent information, both its source and its arrival. We conclude with the issues of comparing and combining information sources. In a subsequent chapter, once our skills are well-practiced, we will extend the earlier treatment of accounting stock and flow measures to an uncertain world. Then we will be in a position to study the information content perspective, and our claim accounting uses the language and algebra of valuation to convey information.

A word of caution: being serious about information is no easy task. Developing the skill to understand and be facile with information issues requires patience and effort. The modeling device we emphasize for this purpose, that of a partition, is likely to appear excessively formal, if not needlessly awkward, at first blush, but it will pay considerable dividends in subsequent chapters.

Modeling Uncertainty

Uncertainty is all around us: What will tomorrow's weather be? When will my hard disk

[1] Ironically, we teach accounting by stressing certainty: here is a set of events and here are the proper procedures; now practice until you can reproduce the correct answer.

fail? Will the driver in the other lane swerve as I attempt to pass? Is there a material error in this asset account? Take your pick. We are also familiar with the use of probability in describing the uncertainty. For example, the probability of precipitation tomorrow is 40%. The probability of heads on the toss of a fair coin is .5.

state-act-outcome specification

It will, in fact, be useful to be formal, almost pedantic about modeling uncertainty. We follow Savage [1954] and model a resource allocation exercise or decision in terms of states, acts, and outcomes. An act is a specific choice, for example attend some specific movie, invest in a specific combination of financial instruments, or launch a specific new product. The act is distinctly endogenous. An outcome is whatever of consequence follows from the choice of act, enjoyment of the movie, profit from the investment, success from the new product (including not only profit but reputation and self-fulfillment).

Certainty, of course, is the case where an act completely determines the outcome. Typically, though, we do not know the precise outcome that will follow from all available acts. This is where states enter. A state is a description of the world so comprehensive that if we know the state, we then know the outcome that will follow from any act.

Let S denote the set of possible or conceivable states, and $s \in S$ a particular state. Likewise, let X denote the set of possible outcomes, and $x \in X$ one such outcome. A, in turn, is the set of possible acts and $a \in A$ is one such act.[2] From here we describe the connection among states, acts and outcomes with a function $x = p(s,a)$. The function $p(s,a)$ catalogues the outcome that will follow if act $a \in A$ is chosen and state $s \in S$ "obtains."[3]

To illustrate, suppose an individual is planning a walk. The remaining issue is whether to carry an umbrella; so the choices are to carry one or not, so $A = \{$*no umbrella, umbrella*$\}$. The individual only cares about whether he gets wet during the walk. This suggests the outcomes are "wet" or "dry." So $X = \{$*get wet, stay dry*$\}$. Notice there is no reason whatsoever to insist the outcomes be monetary.

Also, in this setting the set S would contain two events: it rains, s_1, or it does not rain, s_2. $S = \{s_1, s_2\}$. This leaves no important aspect of the decision problem out of the state description. The outcome in this example might then be described by the matrix, or $p(s,a)$ function, in Exhibit 1. Of course if the decision maker is an airline pilot, this description of the state of the world in terms of rain or no rain is far from sufficient to describe the outcome of following one flight plan versus another.[4]

---

[2]In subsequent chapters we will employ a more mnemonic description of the acts.

[3]For a state to "obtain" is a hopelessly formal description. But it nicely connotes the idea of the realization of some exogenous variable.

[4]Were we insisting on purity, we would admit to a single specification of the set of states, and then any specific decision problem would allow us to group states together, into events, whenever distinguishing among some finely detailed states was immaterial. For example, "rain a little" and "rain a lot" are grouped together in state s_1 in Exhibit 1.

	s_1	s_2
a_1 = no umbrella	*get wet*	*stay dry*
a_2 = umbrella	*stay dry*	*stay dry*

Exhibit 1: State, Act, Outcome Setup for Casual Stroll

For a second illustration, one closer to our explicit concerns, suppose a decision maker is facing two alternatives. Incremental cash flow is the important outcome. One act will result in a zero incremental cash flow. The other will result in a negative incremental cash flow of 100 or a positive incremental cash flow of 140. Then the complete description of the problem is contained in the matrix, or $p(s,a)$ function, below.

	s_1	s_2
a_1	0	0
a_2	-100	140

Exhibit 2: Setup for Risky Investment

For a third illustration, suppose a decision maker faces two investment alternatives, each lasting two periods. Incremental cash flow is again the important outcome, but now this incremental cash flow outcome is an issue in each of the periods. One of the alternatives will result in a zero outcome in both periods. The other will provide incremental cash flows as given below, where the first number in the ordered pair is the incremental cash flow in the first period, and the second is its counterpart in the second period.

	s_1	s_2	s_3	s_4
a_1	0	0	0	0
a_2	(-100,100)	(-100,120)	(-140,130)	(-140,160)

Exhibit 3: Setup for Two Period Risky Investment

As a final example, return to our model of the reporting entity where capital (K) and labor (L) are combined to produce output (q) via the technology expression $q \le \sqrt{KL}$. Further suppose the firm must supply either $q = 100$ or $q = 200$ units of output. The customer will announce the required amount after capital, K, has been acquired, but before labor, L, has been acquired. So we basically have two states, high or low demand for output, and for some initial (and feasible) capital choice of K, the necessary labor to produce q units is the solution to $q = \sqrt{KL}$, or $L = q^2/K$. Also recall the price of capital is p_K per unit and the price of labor is p_L per unit. With this in mind, the total expenditures to meet the customer's requirements, given initial choice of K, would be as described in Exhibit 4.

	s_1 **(q = 100)**	s_2 **(q = 200)**
K	$p_K K + p_L(100)^2/K$	$p_K K + p_L(200)^2/K$

Exhibit 4: Setup for Uncertain Customer Demand

Some common features of these illustrations should be noted. First, we will typically treat the states as finite in number. This means we can write the set of possible states as $S = \{s_1, s_2, ..., s_n\}$ for some number n. This simplifies what follows, and does not cause any significant drop in insight. Second, the relevant description of states depends on available acts and important consequences. Rain or not, for instance, was adequate for the umbrella description, but not for the pilot.[5] Third, each state represents one possible scenario that might occur. All conceivable scenarios are condensed in the state descriptions leaving out no aspects of importance. The notion of states is abstract and can be used in all circumstances where we do not know everything the future might bring (or for that matter that is hidden in the past).

probability

It is also natural to attach the term likelihood to the states. Some states are more likely to occur than other states, just as others may be equally likely. For example, cold weather is more likely than hot weather in the Northern hemisphere during the month of January. The usual way of encoding this into an abstract description of an uncertain world is to attach probabilities to each of the states contained in S.

Recall with a finite set of states we write $S = \{s_1, s_2, ..., s_n\}$. In simplest terms, we think of probability in this context as the assignment of a number π_j for each $j = 1, ..., n$ such that:

(1) $\pi_j \geq 0$ for $j = 1, ..., n$; and

(2) $\sum_{j=1}^{n} \pi_j = 1.$

The number π_j is the probability that state s_j obtains, and will generally be written as *prob*(s_j). The probability assignments, the numbers, are non-negative, and sum to unity.[6]

---

[5]In a deeper sense, once again we are dealing with a single specification of the set of states, and then grouping irrelevant distinctions together, into events or into what you will learn to call a partition of the set of states. This grouping phenomenon was coined the "payoff relevant description of events" by Marschak, an important contributor to the development of the theory of information.

[6]To add some perspective, we dig a bit further. Any subset of S is called an *event*. Thus, $\subseteq S$ is an event, and this holds for any subset of S. Now let be the set of all events, the set of all subsets of S. For example, if $S = \{s_1, s_2\}$, the set of all events is $= \{\emptyset, \{s_1\}, \{s_2\}, \{s_1, s_2\}\}$. Yes, the set of all events is a set; and S itself is an event, just as the null set, $\emptyset$, the null event, is an event. Now, a probability measure on S is a real-valued function π defined on all subsets of S (i.e., defined on all events) such that:

(1) $\pi(\mathscr{E}) \geq 0$ for every $\mathscr{E} \subseteq S$;

(2) $\pi(S) = 1$; and

(3) $\pi(\mathscr{E}_1 \cup \mathscr{E}_2) = \pi(\mathscr{E}_1) + \pi(\mathscr{E}_2)$ when $\mathscr{E}_1, \mathscr{E}_2 \subseteq S$ and $\mathscr{E}_1 \cap \mathscr{E}_2 = \emptyset$.

So, at this level we speak of the probability of an event, the probability of a union of events, and so on. The probability is always non-negative and has a casual interpretation as the degree of likelihood. $\pi(\mathscr{E}_1) > \pi(\mathscr{E}_2)$ means event $\mathscr{E}_1$ is more likely

A remaining question before we proceed is the source of this probability measure. We will treat it as exogenous, it is just assumed to be present. The very meaning one might attach to the probability measure has been the subject of philosophical argument, for example is this a long run frequency notion, a completely subjective notion, or perhaps somewhere in between? Savage [1954] pioneered a subjective interpretation, where a person whose behavior in decision making is sufficiently consistent can be modeled as if subjective probability and subjective utility assessments are made and decision theory calculus is invoked to identify the most preferred act. We treat the probability as a measure of likelihood, where that likelihood might reflect "objective" events such as gambling encounters with known odds or "subjective" events such as new product introductions with anything but objective odds.[7]

This construct of probabilities assigned to state descriptions, then, captures the notion of the uncertainty. If we know which state will occur, we assign a probability of one to that state and a probability of zero to all other states. If we find a state very likely to occur, we assign a probability close to one to that state and reversely if a state is very unlikely, the probability assigned to that state will be close to zero.

random variables

States and probabilities, then, are our foundation for modeling uncertainty. In most cases, though, the state description is of little or no practical use. (No comment is necessary at this point.) We don't speak of the state of the weather, but rather in terms of average temperatures, temperature range, precipitation expected, and so on. Similarly, we don't speak of the state of the enterprise, but instead in terms of its earnings, growth prospects, book to market ratio, and so on.

The connection between this foundation and the things we so casually speak of is the concept of a random variable. Given state description $S = \{s_1, s_2, ..., s_n\}$, a random variable is a numerical

---

than event $\mathscr{E}_2$. (Parenthetically, it is also important to note our reliance on a finite *S* here. If *S* is a richer set, we must be careful to specify the collection of subsets of *S*, their complements, and so on. This leads us into things like a σ-algebra, which is a set and collection of subsets thereof where each subset's complement is present, and where the union of a countable number of subsets is also a subset. With a finite *S*, though, we work with all subsets of *S*, without causing any ambiguities or measurement difficulties.)

Now $\{s_j\}\subseteq S$ is the event "state s_j" and $\{s_i, s_j\}\subseteq S$ is the event "state s_i or state s_j." (Don't miss the subtlety: s_j is an element of set *S*, but an event is a subset of the set *S*. Properly speaking, then, $\{s_j\}\subseteq S$ is the event "state s_j" and when we write the probability of state s_j as $prob(s_j)$ we should be writing $prob(\{s_j\})$.) And since the events "state s_i" and "state s_j" are distinct events, given $i \neq j$, the probability of any event in our setup is the sum of the "state probabilities" for each state contained in the event: $\pi(\mathscr{E}) = \sum_{s\in} prob(s)$.

[7]Probability is a measure, a measure of likelihood. Suppose we have a set of events, as in our finite state setup. Further suppose we are able to rank all of the events in terms of likelihood. Let this ranking satisfy the following four properties: (1) it is complete and transitive; (2) any event is at least as likely as the null event; (3) *S* itself is strictly more likely than the null event: and (4) if $\mathscr{E}_1$ and $\mathscr{E}_3$ are disjoint events and if $\mathscr{E}_2$ and $\mathscr{E}_3$ are disjoint events, then $\mathscr{E}_1$ is strictly more likely than $\mathscr{E}_2$ if and only if $\mathscr{E}_1\cup\mathscr{E}_3$ is strictly more likely than $\mathscr{E}_2\cup\mathscr{E}_3$. Our noted definition of a probability measure satisfies all of these conditions. But is the opposite true? This was an open question for a long time, but the answer is no. (See Kraft, Pratt, and Seidenberg [1959].) The point is that treating probability as a measure of likelihood is both natural and delicate. It requires considerable structure on what we mean by likelihood for this representation to be valid. Accounting theory is likewise concerned with measurement: what, how and when should the accounting system measure? The idea, you will come to appreciate, is to use the measures to represent the underlying set of phenomena.

valued function which assigns a real number $f(s)$ to every state $s \in S$. Technically, a random variable is a function which is defined on the state space and maps into the real line (perhaps an n-dimensional space).

Rainfall, temperature, accounting income, and your score on the final exam are all random variables. Likewise, in the state-act-outcome setup, if the outcome is a real number, such as cash flow, it too is a random variable.

To illustrate, consider the setting described in Exhibit 5, where we have five possible states, and two different random variables. Z is a random variable, defined by the function $f(s)$. So is G, as defined by the function $d(s)$.

	s_1	s_2	s_3	s_4	s_5
$\pi_j = prob(s_j)$	.2	.2	.3	.2	.1
$Z = f(s)$	1	2	2	4	-1
$G = d(s)$	5	5	6	5	6

Exhibit 5: Illustrative Random Variables

Notice the probability measure is also specified.[8] This probability measure combined with the function, $f(s)$, allows us to identify a probability measure for the random variable (technically for the events associated with the random variable). Thus, $prob(Z = 1) = .2$, $prob(Z = 2) = .2 + .3 = .5$, etc. We summarize the details in the following matrix.

	$Z = -1$	$Z = 1$	$Z = 2$	$Z = 4$
$prob(z)$	.1	.2	.5	.2

Exhibit 6: *prob*(*z*) specification

Clearly this leads to a probability measure defined on all subsets or events of the set {-1, 1, 2, 4}!

From here it is a short step to start talking about summary measures, such as the mean or expected value and variance of a random variable. For our random variable Z we readily calculate the mean, $E[Z]$, and variance, $VAR(Z)$:

$$E[Z] = \sum_z z \cdot prob(z) = -1(.1) + 1(.2) + 2(.5) + 4(.2) = 1.9; \text{ and}$$

$$VAR(Z) = \sum_z (z - E[Z])^2 \cdot prob(z) = (-1-1.9)^2(.1) + (1-1.9)^2(.2) + (2-1.9)^2(.5) + (4-1.9)^2(.2) = 1.89.$$

The calculations, you will appreciate, rely on the fact Z is numerical.

A more familiar specification is when we directly assess the random variable's probability measure. For example, the income of a firm for the forthcoming year is a random variable. It

---

[8]You should have no difficulty moving from the given $prob(s)$ specification to the probability of any event $\mathscr{E} \subseteq S = \{s_1, s_2, ..., s_5\}$. For example, the probability of "s_1 or s_2," i.e., the probability of the event $\{s_1, s_2\}$ is simply $\pi_1 + \pi_2 = .4$.

might be any conceivable real number. It might also be well described by a random variable, $\hat{I}$, which is normally distributed with mean $E[\hat{I}] = \mu$ and variance $VAR(\hat{I}) = \sigma^2$ (which, for the notation connoisseur, is usually written as $\hat{I} \sim N(\mu,\sigma^2)$). The two parameters, mean μ and variance σ^2, are sufficient to describe the entire distribution.[9]

The final step is to introduce the notion of joint probabilities of the random variables Z and G. This is the probability that the pair Z, G takes on a specific value. For example $prob((Z,G) = (1,5)) = .2$. The notion of joint probability conveys the dependency among the random variables. This becomes important once we skip the state space specification, because the states that carry all uncertainty also keep track of the interrelationship among all random variables. In the random variable formulation this is replaced by a joint probability measure. Covariance is a well known summary measure of this dependency.

Information

Given the state description contains every conceivable relevant aspect of the world it is especially simple to define the notion of information. At one extreme perfect information is equivalent to learning exactly which state will occur. At the other extreme is null information, where we learn nothing about which state will occur. The intermediate case is where we learn something, but less than all there is to know.

To model this, recall the state description in the state-act-outcome setup carries, so to speak, the uncertainty in the resource allocation exercise. Intuitively, then, information should "refine our knowledge" of the states.

For example, suppose we have four states, $S = \{s_1, s_2, s_3, s_4\}$. Further suppose an information source will reveal something about these states. In particular, it will report one of two possible reports: the state that will obtain, the true state, is either a member of $\{s_1, s_2\} = \delta_1$ or the state that will obtain is a member of $\{s_3, s_4\} = \delta_2$. So the report will be either δ_1 or δ_2. Details are given below, including $prob(s)$.

	s_1	s_2	s_3	s_4
$\pi_j = prob(s_j)$	.1	.2	.3	.4
$prob(s \mid \delta_1)$	1/3	2/3	0	0
$prob(s \mid \delta_2)$	0	0	3/7	4/7

Exhibit 7: Probability Details for Information Example

Now, what do we learn if this source reports δ_1? We learn the true state is either s_1 or s_2 or formally that the state is a member of $\delta_1 = \{s_1, s_2\}$, and therefore not a member of $\{s_3, s_4\} = \delta_2$.

[9]The Normal density is completely described by the two parameters of mean and variance. In this example, now, income $\hat{I}$ can take on any value, so we are clearly outside our usual setup of a finite set of states. Our earlier warning about technical issues now resurfaces. We know the mean and variance of a Normally distributed random variable exist. In general, though, the event algebra has to be carefully specified to ensure, say, the mean as a mathematical construction exists. This is far beyond any technical measurability issue we will encounter; but you should be aware we are not presenting a comprehensive treatment of the subject.

Surely, then, the probability of either of the latter two states, given we have observed δ_1 is zero. We write this in conditional probability format as $prob(s_3 \mid \delta_1) = prob(s_4 \mid \delta_1) = 0$. What about the other two states? Here we calculate

$$prob(s_1 \mid \delta_1) = \pi_1/(\pi_1 + \pi_2) = .1/(.1 + .2) = 1/3; \text{ and}$$
$$prob(s_2 \mid \delta_1) = \pi_2/(\pi_1 + \pi_2) = .2/(.1 + .2) = 2/3.$$

Note well. Observing report δ_1, or report δ_2 tells us something about the states. This, in turn, leads to systematic revision of the probabilities. Begin with the noted probability assessment, $prob(s)$. If we subsequently learn the source reports $\delta_1 = \{s_1, s_2\}$ (or $\delta_2 = \{s_3, s_4\}$), we use Bayes' Rule to revise $prob(s)$ to the associated conditional probability, conditional on having observed report δ:[10]

$$prob(s \mid \delta) = \frac{prob(\{s\} \cap \delta)}{prob(\delta)}.$$

You should verify the calculations for $prob(s \mid \delta_1)$ and $prob(s \mid \delta_2)$.

partitions

An additional feature of the above example to notice is that each possible report is actually a subset of the set of states, S. Moreover, the two possible subsets, δ_1 and δ_2, share nothing in common, i.e., $\delta_1 \cap \delta_2 = \emptyset$, and collectively define S, i.e., $\delta_1 \cup \delta_2 = S$. In formal terms, δ_1 and δ_2 form a partition of the set S. This is a consequence of the comprehensiveness of the state description. It includes all conceivable uncertainty, and that of course includes any information we might receive.

To illustrate, suppose the weather might be "wet" or "dry," and a weather forecast will predict "wet" or "dry." Naturally, this forecast might turn out to be correct, or erroneous. We codes this as four states: S = {*wet and forecast wet, wet and forecast dry, dry and forecast wet, dry and forecast dry*}. So a forecast of wet is a claim the first or the third state is true, etc. In this fashion the state specification tautologically reflects all uncertainties, including those associated with the implications of what some information source reports. As a consequence, whatever that information source is, it defines a partition on the state space.

To go a bit further, let $\Delta = \{\delta_1, \delta_2, ..., \delta_m\}$ be a collection of sets. This collection of sets, Δ, forms a partition of $S = \{s_1, s_2, ..., s_n\}$ if:

---

[10] A more familiar expression for Bayes' Rule relies on random variables. Suppose Z and W are random variables, with joint probability $prob(Z,W)$. Having observed $W = w$, the probability that $Z = z$ is given by

$$prob(Z = z \mid W = w) = prob(Z = z, W = w)/prob(W = w),$$

or in our shorthand notation, $prob(z \mid w) = prob(z, w)/prob(w)$. In our case, though, the probabilities are defined on subsets of S, and δ itself is a subset of S; so the joint occurrence of s and δ is the intersection: $\{s\} \cap \delta$. Likewise, $prob(\delta)$ is simply the sum of the underlying state probabilities for each state contained in δ. In particular, $\{s\} \cap \delta_1 = \{s\}$ for the first two states, and is null for the last two. So, $prob(\{s_1\} \cap \delta_1) = prob(s_1) = .1$, along with $prob(\{s_2\} \cap \delta_1) = prob(s_2) = .2$; $prob(\{s_3\} \cap \delta_1) = prob(\emptyset) = 0$; and $prob(\{s_4\} \cap \delta_1) = prob(\emptyset) = 0$. In addition, with $\delta_1 = \{s_1, s_2\}$, $prob(\delta_1) = prob(s_1) + prob(s_2) = .1 + .2 = .3$. This provides $prob(s_1 \mid \delta_1) = .1/.3 = 1/3$, and so on.

(1) $\delta_i \subseteq S$, for every $\delta_i \in \Delta$;
(2) $\delta_1 \cup \delta_2 \cup \cdots \cup \delta_{m-1} \cup \delta_m = S$; and
(3) $\delta_i \cap \delta_j = \varnothing$ for every δ_i, $\delta_j \in \Delta$ and $i \neq j$.

Think of Δ as defining a set of "holders." First, each "holder," each δ, must be a subset of S. Holder δ is not allowed to have any elements outside of S. Second, collectively, the "holders" must equal S. Every element of S must be placed in one of the holders. Finally, no ambiguity is allowed; the "holders" are mutually exclusive. The idea is classification: every element of S belongs to exactly one element of Δ.

This is how we model information. It is a partition of the state space S. Suppose Δ contains a single set. This single set must be S itself, i.e., $\Delta = \{S\}$. Otherwise we have not satisfied the definition of a partition. This is null information, it tells us nothing. After all, we already know $s \in S$! At the other extreme, suppose $\Delta = \{\{s_1\},\{s_2\},...,\{s_n\}\}$. This is perfect information.

Two additional features of this modeling apparatus are important. First, as we observed earlier, revision of the state probabilities is particularly straightforward when information is modeled as a partition of the state space. If we know $\delta \in \Delta$ is true, we then know the true state is one of the elements of δ, and by implication any state not in δ has now been ruled out. So, our earlier, surely awkward, statement of Bayes' Rule simplifies to

$$prob(s \mid \delta) = \begin{cases} prob(s)/prob(\delta) \text{ if } s \in \delta\text{; and} \\ \\ 0 \text{ if } s \quad \delta. \end{cases}$$

Second, an equivalent way to think about this construction of information as a partition of the state space is that the information source reports according to some function η that maps S into some set of possible signals, denoted . We formally state this as η: $S \rightarrow$, meaning "η maps S into ." Alternatively, we write this as $m = \eta(s)$.[11] In this construction we identify the information source, η, and the signal or message, $m \in$ it provides. Formal equivalence between the function and partition ideas should be evident by glancing back at our earlier example.

Indeed, we could go further here and think of this in terms of a random variable. Instead of

---

[11]The Greek letter η ("eta") is the traditional symbol for an information structure defined in this manner, just as ϵ is the traditional symbol for the error term in a regression equation, μ is the traditional symbol for the mean of a normal population, etc. Now, recall, a function is a mapping from one set to another with two properties: no element of the first set is left "unmapped" and no element of the first set is mapped into more than one element of the second set. Thus, if $\eta(s)$ is a function from S, it defines a partition of S. Suppose we observe $m = \eta(s)$. Then the inverse, $\eta^{-1}(m)$, identifies all elements of S that lead to m.

Treating information as a partition of S, then, is equivalent to treating the information source as providing a signal defined by the function $m = \eta(s)$. Subsequently, we will worry about information that is useful in monitoring an agent, an agent who selects the act a on behalf of someone else. We will then treat the monitor as reporting a signal defined by $\eta(s,a)$; that is, we will allow the agent's behavior to specify which partition of S we are observing. Stated differently, the information source will partition $S \times A$ in that setting. In similar fashion, an accounting procedure will take various activities of the entity (viewed as entity acts) and compile an accounting rendering, based on other available information.

reporting an element of the partition Δ, i.e., instead of reporting $\delta_j \in \Delta$, why not simply report the number "*j*." Go back to our example. $\eta(s_1) = \eta(s_2) = 1$ and $\eta(s_3) = \eta(s_4) = 2$ defines a random variable and this random variable conveys the same information, the same underlying partition, as was used in the illustration. The substance of the information, though, is not whether the random variable's realization was 1 or 2, it is the underlying state partition, the underlying information content. For example, the substance of a firm reporting income of so many dollars is not that this amount of income was earned, or recognized. It is what you learn about the firm, given its accounting system has reported this income number.

alternative representations

Our partition (or function) formulation, though awkward at first encounter, is the simplest device on which to base our forthcoming study of accounting as a source of information. Its (full) generality can be appreciated by briefly considering two alternative ways of representing the arrival of information. For this purpose, suppose we are interested in a random variable $D \in \{0, 1\}$. We also have access to an information source that will report "good" (*g*) or "bad" (*b*) news. The joint probability is specified as follows.

	***D* = 0**	***D* = 1**
signal *g*	.15	.40
signal *b*	.35	.10

Exhibit 8: Binary Random Variable

Notice $prob(D = 0) = prob(D = 1) = .5$, and $E[D] = 0(.5) + 1(.5) = .5 = prob(D = 1)$. In addition, $prob(\text{signal } g) = .55$ and $prob(\text{signal } b) = .45$. Observing the noted signal is informative, it alters our opinion about the variable *D*. This becomes more apparent when we calculate the expected value[12] of *D*, conditional on either of the signals:

$$E[D \,|\, \text{signal } g] = 0(15/55) + 1(40/55) = 40/55;$$
$$E[D \,|\, \text{signal } b] = 0(35/45) + 1(10/45) = 10/45.$$

Can you represent this story in terms of a state space and the information itself as a partition Δ of this state space? See Exhibit 9, where we merely enrich the state description so it captures all the noise in the information source as well.

Notice $\delta_1 = \{s_1, s_3\}$ corresponds to signal *g*; and we readily have $prob(s_1 \,|\, \delta_1) = .15/(.15 + .40) = 15/55$ and $prob(s_3 \,|\, \delta_1) = 40/55$. Clearly, $E[D \,|\, \delta_1] = 40/55$. The remaining details should be transparent. We are telling the identical story.

Intuitively, this should be the case. After all, the state is specified so it carries all the uncertainty. If there is uncertainty between the variable *D* and the signal, as in Exhibit 5, we merely "load" all of that uncertainty into an equivalent state specification.

[12] Again notice that because *D* can only be 0 or 1 its expected value is numerically equal to the probability that $D = 1$.

	s_1: D = 0 and signal = g	**s_2: D = 0 and signal = b**	**s_3: D = 1 and signal = g**	**s_4: D = 1 and signal = b**
***prob*(*s*)**	.15	.35	.40	.10
$\eta(s)$	δ_1	δ_2	δ_1	δ_2
$prob(s \mid \delta_1)$	15/55	0	40/55	0
$prob(s \mid \delta_2)$	0	35/45	0	10/45

Exhibit 9: Partition Version of Exhibit 8 Setting

Yet another way to model this story is to focus directly on the random variable D. Suppose all we care about is the expected value of D. Well, absent any information, this is simply .50. But, if signal g (or partition element δ_1) is observed, we know the expected value increases to 40/55; and if signal b (or partition element δ_2) is observed, we know the expected value decreases to 10/45. So, let's represent the expected value as the random variable $\overline{D}$:

$$\overline{D} = .5 + \epsilon.$$

ϵ, now, is a zero mean shock or disturbance term. Try the following: ϵ = 125/550 (which is 40/55 - .50) with probability .55 and ϵ = -125/450 (which is 10/45 - .50) with probability .45. So

$$E[\overline{D}] = .5 + E[\epsilon] = .5 + .55(125/550) + .45(-125/450) = .5.$$

That is, the revised expected value of D can be modeled as equal to its mean plus a random "innovation." In information terms, then, we observe the innovation or revision in the mean. Stated differently, information alters the expectation in this case.

Thus, we generally have alternative ways of modeling or representing information. The joint probability representation has the advantage of familiarity, but it becomes awkward when we compare information sources (as you will see). The innovation representation is intuitive, but it becomes awkward if not dysfunctional when we have multiple sources of information. The partition approach has the advantage of readily accommodating multiple sources of information and being particularly transparent on the subject of comparing information sources.

At times we will switch among these representations for expositional ease. But the more subtle issues will always lead us back to the partition formulation.

Comparison Of Information Sources

Next is the question of whether there is some way to order or compare information sources in terms of their "usefulness" or "value." For example, is it possible to say one newspaper is better than another, or to say one accounting procedure is better than another? To explore this, let S have three elements $S = \{s_1, s_2, s_3\}$. Below we list all possible partitions of S.

label	partition	interpretation
Δ_0	$\{\{s_1, s_2, s_3\}\}$	tells us nothing

$\boldsymbol{\Delta_1}$	$\{\{s_1\},\{s_2, s_3\}\}$	highlights s_1
$\boldsymbol{\Delta_2}$	$\{\{s_2\},\{s_1, s_3\}\}$	highlights s_2
$\boldsymbol{\Delta_3}$	$\{\{s_3\},\{s_1, s_2\}\}$	highlights s_3
$\boldsymbol{\Delta_4}$	$\{\{s_1\},\{s_2\},\{s_3\}\}$	tells us everything

Exhibit 10: Partitions of a Three Element *S*

Partition Δ_0 is of course null, it tells us nothing; and partition Δ_4 is perfect, it tells us exactly which state will obtain. The other three are in between, each distinguishes one of the three states, and groups the other two.

Now suppose we are given, say, partition Δ_1. Could we convert this into partitions Δ_4 by "splitting apart" or "subdividing" one or more of its elements? The answer is yes; simply take the second element of Δ_1, $\{s_2, s_3\}$, and split it into $\{s_2\}$ and $\{s_3\}$. This procedure goes by the name subpartition. We take a given partition and "subdivide" it. Disaggregation is an apt metaphor.

Here is the formal idea. Suppose $\Delta = \{\delta_1, ..., \delta_m\}$ and $\underline{\underline{\Delta}} = \{\underline{\underline{\delta}}_1, ..., \underline{\underline{\delta}}_{m'}\}$ both partition state space *S*. Partition Δ is a subpartition of partition $\underline{\underline{\Delta}}$ if for every $\delta \in \Delta$ there exists a $\underline{\underline{\delta}} \in \underline{\underline{\Delta}}$ such that $\delta \subseteq \underline{\underline{\delta}}$. In words, if we can take any element of partition Δ and find a corresponding element in partition $\underline{\underline{\Delta}}$ that contains that element, then Δ is a subpartition of $\underline{\underline{\Delta}}$.

Try this on our three state example.

label	partition	subpartition of
$\boldsymbol{\Delta_0}$	$\{\{s_1, s_2, s_3\}\}$	nothing but Δ_0
$\boldsymbol{\Delta_1}$	$\{\{s_1\},\{s_2, s_3\}\}$	Δ_0, Δ_1
$\boldsymbol{\Delta_2}$	$\{\{s_2\},\{s_1, s_3\}\}$	Δ_0, Δ_2
$\boldsymbol{\Delta_3}$	$\{\{s_3\},\{s_1, s_2\}\}$	Δ_0, Δ_3
$\boldsymbol{\Delta_4}$	$\{\{s_1\},\{s_2\},\{s_3\}\}$	Δ_0, Δ_1, Δ_2, Δ_3, Δ_4

Exhibit 11: Subpartition Relationships

Partition Δ_0 can be subpartitioned to create any of the other partitions. Δ_4, being perfect information, is a subpartition of every other partition but cannot itself be subpartitioned. Δ_1 is not a subpartition of Δ_2 and vice versa. And, naturally, any partition is a subpartition of itself.

Think of a subpartition as providing more detail. If Δ is a subpartition of $\underline{\underline{\Delta}}$, then anything $\underline{\underline{\Delta}}$ might tell you will also be revealed by partition Δ, along with possibly additional details. For example, consider Δ_1 and Δ_4 in Exhibit 11. The latter is simply the former, but with subset $\{s_2, s_3\}$ subdivided into subsets $\{s_1\}$ and $\{s_2\}$. That is, Δ_4 is a subpartition of Δ_1. Another way to see this is to begin with Δ_4. Now notice we can now construct partition Δ_1 by combining subsets $\{s_2\}$ and $\{s_3\}$ into subset $\{s_2, s_3\}$.

A synonym for subpartition is "as fine as" (or, more emphatically, at least as fine as). Try it out. Δ_1 is as fine as Δ_0. Δ_4 is as fine as Δ_1 but not vice versa. Δ_1 is not as fine as Δ_2 and Δ_2 is not as fine as Δ_1. Δ_4 is the finest partition of *S*. Δ_0 is the least fine, or coarsest.

Two features of this odyssey should be noted. First, suppose we tell you partition Δ is a subpartition of partition $\underline{\underline{\Delta}}$, or equivalently partition Δ is as fine as partition $\underline{\underline{\Delta}}$. Then everything you might learn from partition $\underline{\underline{\Delta}}$ you can learn from partition Δ, and possibly more. Thus, partition Δ provides as much information as does partition $\underline{\underline{\Delta}}$. In this limited sense we can rank the information sources. If one is a subpartition of the other, we know it provides at least as much information as the second.

Second, this subpartition or fineness device provides a partial but not a complete ranking of partitions, or information sources. Compare partitions Δ_4 and Δ_3 in our three state example. Partition Δ_4 is a subpartition of partition Δ_3. Δ_4 is as fine as Δ_3, but not vice versa. Now try to compare partitions Δ_2 and Δ_3. Neither is a subpartition of the other. They cannot be compared via the subpartition or fineness device. The implication should not be missed: it is simply not always possible to compare two partitions or information sources and claim that one information system is superior to another, in terms of providing more information.[13]

This fact turns out to be important in a variety ways, so we amplify a bit. Consider an ordering of a set of objects by some criterion, e.g., individuals by height or partitions of some state set using the subpartition idea. Now, this ordering is complete if for any two objects in the set, one is ordered above the other or vice versa. Completeness means we can always compare the two objects using the noted criterion. Similarly, the ordering is transitive if when one object is ordered above a second and the second is ordered above a third, then the first is ordered above the third. We will only call this ordering a ranking if it is both complete and transitive. Formally these are the characteristics we combine with ranking. The first makes sure that all objects are ranked and the second rules out circularity. Only then can we talk about the highest and the lowest ranked object.

The punch line is the subpartition criterion provides a ranking that is transitive, but not complete. Consequently we only find a partial ranking of information sources using the subpartition or fineness criterion.

equivalent information sources

Closely related to the notion of subpartition is the notion of equivalent information sources. Partitions Δ and $\underline{\underline{\Delta}}$ are equivalent if they are identical, if $\Delta = \underline{\underline{\Delta}}$. Equivalent information sources tell us the same thing. This amounts to (1) Δ is a subpartition of $\underline{\underline{\Delta}}$ and (2) $\underline{\underline{\Delta}}$ is a subpartition of Δ.

As obvious as this is, we should remember information typically does not arrive in the form of an explicit state partition element. It is generally coded. Think of an important strategic report, written in both German and English. If nothing was missed in translation, this is the same information but in a different code, scale, or language.

Earlier we noted another way to represent the partition idea is to think in terms of a function

---

[13]In the random variables setup the notion of better information transforms to the condition that one random variable W is more informative than another random variable Z with respect to some random variable B if there exists a random variable ϵ which is independent of B and has $E[\epsilon] = 0$ such that

$$Z = W + \epsilon.$$

The interpretation is that random variable Z is equal to the random variable W plus a noise term which is totally unrelated to the variable of interest, B.

that maps states into some set of admissible signals. For example, the function $\eta(s)$ might report a real number for each of the states. Let's concentrate on systems that do use real numbers to reveal what they know about S. So $\eta(s)$ is a real valued function. (Yes, it defines a random variable.) Now suppose we have two such functions, $\eta_1(s)$ and $\eta_2(s)$. Further suppose for at least one state s_i their reports differ: $\eta_1(s_i) \neq \eta_2(s_i)$. Are these different information systems? They surely are in the sense their reports will not always agree, literally.

But this is naive. Celsius and Fahrenheit scales tell us the same thing, but always with a different temperature reading except at negative 40 degrees. To deal with this we must identify the partitions induced by the information systems in general.

The important aspect of an information system is what it tells about the underlying state space, i.e. what partition of the state space it induces. Thus if the two information structures induce the same partition, the two information structures are equivalent. In that case the two information systems carry the same information about the underlying state space. That will be the case whenever η_2 can be constructed from η_1 and vice versa.[14] In that case there exists functions F and G such that for all states we have $\eta_1(s) = F(\eta_2(s))$ and $\eta_2(s) = G(\eta_1(s))$. Stated differently, in that case the partitions of the state space induced by each of the systems are identical. The two systems have precisely the same information content, but deliver it with a different code, measurement scale, or simply scale.

This equivalence of information structures leads to the observation there might be many equivalent representations of an information source. The only difference among them is the labeling of the partition elements. The face value of an information system is not the source of its substance. It is the induced partition of the state space that matters.

combining information sources

In many cases more than one information structure is available. The mechanics are straightforward. Consider the case where two information systems are jointly available, defined by partitions Δ and $\underline{\underline{\Delta}}$ alone. Taken together, the two partitions provide a partition of the state space that is as fine as either Δ or $\underline{\underline{\Delta}}$. The second source, so to speak, can only improve upon the first source.[15] To illustrate, let $S = \{s_1, s_2, s_3, s_4\}$. Also assume $\Delta = \{\{s_1, s_2\},\{s_3, s_4\}\}$ and $\underline{\underline{\Delta}} = \{\{s_1, s_3\},\{s_2, s_4\}\}$. Combining the two partitions provides us the partition of $\{\{s_1\},\{s_2\},\{s_3\},\{s_4\}\}$.

As we said, the mechanics of combining information sources are straightforward. What is far from straightforward is understanding the importance of one of the information sources. We will learn that studying one source by itself, say Δ, as though it were the only source of information, tells us in general almost nothing about how important Δ might be in the presence of information source $\underline{\underline{\Delta}}$. Mixing or combining sources may make the source in question vastly more or vastly less important. As a hint of things to come, this is why accounting theory cannot treat the accounting function in isolation. We must carry along other sources of information.

---

[14]Let S have four elements $S = \{s_1, s_2, s_3, s_4\}$. Let $\eta_1(s_1) = \eta_1(s_3) = m_1$ and $\eta_1(s_2) = \eta_1(s_4) = m_2$. Let $\eta_2(s_1) = \eta_2(s_3) = m_3$ and $\eta_2(s_2) = \eta_2(s_4) = m_4$. Then the two information structures are equivalent, as they both induce partition $\{\{s_1, s_3\}, \{s_2, s_4\}\}$. In turn, using $F(m_3) = m_1$ and $F(m_4) = m_2$ converts the second system into the first, and so on.

[15]Why, then, can we not have a situation where the second source "destroys" the first?

Summary

Surely accounting is a source of information. But carefully examining this colloquialism requires considerable setup. A major piece of this setup is modeling uncertainty and information. Our approach is to envision a state variable that tautologically carries all important uncertainty in whatever setting we find ourselves. We assign probabilities to the states to capture the notion of likelihood, both when it is objective and when it is subjective. Uncertainty is thus closely related to the calculus of probability.

Information is then modeled as a partition of this set of possible states. Eventually we will learn to view an accounting system as providing a partition of some set of states, and in this view different "sets of books" are nothing other than different partitions.

The partition idea is at once conceptually useful and awkward. So at times, where this awkwardness becomes distracting, we will switch to equivalent, alternative specifications. Fundamentally, though, uncertainty is encoded in the state specification and an information source is a partition of the set of states.

The partition idea also provides a partial comparison or ranking of information sources. This is based on the idea that if one source reports at least all that a second source might report, then that first source provides "more" information than the second. In the state partition setup this simply means the first partition is a subpartition of the second.

Accounting, of course, is not any old source of information. So we have yet to make clear what it means to say accounting is a source of information, and to address its comparative advantage as a source of some "type" of information (read that partition).

Selected References

Using the state idea to model uncertainty has its roots in Savage [1954], as mentioned. Marschak [1963] provides an excellent treatment. Information is treated in a variety of sources, including Baye [1999], Beaver [1998], Scott [1996], Demski and Feltham [1976] and Demski [1980]. Marschak and Miyasawa [1968] and Feltham [1972] provide excellent, more formal treatments.

Key Terms

Uncertainty is the opposite of certainty, a lack of complete foreknowledge, a setting in which there is ambiguity as to what will transpire. This requires a specification of the "ambiguity as to what will transpire." For this we use the state-act-outcome device, where the outcome that will transpire, x, depends on the act chosen, a, and the state, s: $x = p(s,a)$. The state that "will obtain" is unknown, and this ambiguity is described by a *probability* measure on the set of possible states. *Information*, in turn, is something that provides insight into this ambiguity, something that revises the probability measure on the set of possible states. From a modeling perspective, we find it convenient to describe the information source as simply "shrinking" the set of possible states, in effect revealing something about states that are certain to not occur. This relies on an information source being a *partition* of the set of possible states, a collection of mutually disjoint subsets of the set of possible states whose union is the set of possible states. The information source then reveals which partition element contains the true state. In turn one partition is a *subpartition* of a second partition if it "subdivides" the second, meaning each of its elements is a subset of an element of the second. Subpartition or *fineness* provides a partial ranking of information sources. Two information sources, two partitions, are *equivalent* if they reveal the same about the underlying state, if they are identical partitions.

Problems and Exercises

1. Accessing information presupposes uncertainty. Explain.

2. Suppose you travel across time zones but do not reset your watch. Does your watch provide the same information in the new as opposed to the original time zone? Explain. Relate your answer to the notion of equivalent information structures.

3. Ralph's cash flow is uncertain; it will be either 100 or 400. In addition, Ralph will access an information source that will report signal "*g*" or signal "*b*." The joint probability of cash flow and signal are specified in the following table.

	***CF* = 100**	***CF* = 400**
signal *g*	.10	.50
signal *b*	.30	.10

   Does the information source inform about Ralph's cash flow prospects? Carefully explain your answer. Next, calculate the mean and variance of Ralph's cash flow, conditional on each signal. Finally, provide an equivalent state-based description of this story. Be certain to identify the probabilities as well as the partition provided by the information source.

4. Consider the following setting where *D* might be 10 or 20, and some information source might report "*g*" or "*b*:"

	signal *g*	**signal *b***
***D* = 10**	.10	.40
***D* = 20**	.40	.10

   Determine the expected value of *D*, the expected value of *D* given signal *g*, and the expected value of *D* given signal *b*. What is the expected value of the conditional expected value of *D*? Now, instead of revising the expected value of *D* in this fashion, suppose you think of the revised expected value of *D* as

$$\overline{D} = 15 + \epsilon.$$

Specify the random variable ϵ: what values can it take on, and the probability of each? What is the expected value of ϵ?

Finally, provide an equivalent state-based description of this story, including the information partition.

5. Suppose there are 10 possible, equally likely states: $S = \{s_1, ..., s_{10}\}$. Information source one will report signal "1" if the true state is one of the first five and "2" if it is one of the last five. Source two will report "odd" if the state index of the true state is an odd number (i.e., if the true state is one of s_1, s_3, etc.) and "even" if the state index of the true state is an even number. Specify the partitions of S that are provided by each source. Then specify the partition provided by having access to both information sources. Finally, in the latter case suppose we have a single information source that maps S into the Greek letters $\{\alpha, \beta, \gamma, \delta\}$. Provide two such mappings that are informationally equivalent to observing the original sources one and two.

6. Let $S = \{s_1, s_2, s_3, s_4, s_5\}$. Provide three partitions, one that is a subpartition of the second which, in turn, is a subpartition of the third. Is the first a subpartition of the third? Finally, provide two partitions, neither of which is a subpartition of the other.

7. Repeat Exercise 4 above, using the following probability specification:

	signal *g*	**signal *b***
***D* = 10**	.10	.40
***D* = 20**	.35	.15

8. We now find Ralph thinking about multiple sources of information. Consider the following matrix, concerning the cash flow random variable of *CF* = 100 or *CF* = 200, along with two information sources. One will report "$\mathtt{g}$" or "$\mathtt{b}$" while the other will report "*g*" or "*b*." Probabilities are given in the following table.

	***CF* = 100**		***CF* = 200**	
	g	***b***	***g***	***b***
$\mathtt{g}$	.02	.16	.08	.04
$\mathtt{b}$	.40	.02	.20	.08

Now suppose we can observe both information sources. Determine the expected value of *CF* conditional on no information, on $\mathtt{g}$ only, on $\mathtt{b}$ only, on *g* only, on *b* only, and on $\mathtt{g}$ and *g*, $\mathtt{g}$ and *b*, $\mathtt{b}$ and *g*, and $\mathtt{b}$ and *b*. Also, can you express the above changes in the expectation of *CF* as 140 + ϵ_1 + ϵ_2, where the first "shock" term refers to that from the $\mathtt{g}/\mathtt{b}$ observation and the second from the *g*/*b* observation? Explain.

Now suppose *CF* might be 300, 500 or 700. The only events with positive probability are the following four, and each has probability .25:

> $\mathtt{g}$, *g* and *CF* = 700;
> $\mathtt{g}$, *b* and *CF* = 500;
> $\mathtt{b}$, *g* and *CF* = 500; and
> $\mathtt{b}$, *b* and *CF* = 300.

Again, we will track the expected value of *CF* as the $\mathtt{g}/\mathtt{b}$ event is observed and then the *g*/*b* event is observed. Try to express this story as an expected value (of 500) that changes under additive innovations from the $\mathtt{g}/\mathtt{b}$ event and from the *g*/*b* event. Explain.

July 9, 2001, Joel