# 1 Information

'Information', used in the technical sense, is the converse of an uncertainty (an uncertainty decrease is an increase in a state of information). Appropriate measures of information/uncertainty are designated $H$ (equation 1 in Appendix A) and computed in respect of one universe which is believed to be independent of others. In order to carry out the computation the 'universe' in question must be partitioned, by a description, into a set of exclusive and (for the universe) exhaustive alternatives, one of which must obtain and only one of which may obtain.

For example, a universe $U$ may be a choice set, the (set of) values of a state variable or a set of states, determined by the conjoint values of several state variables.[1]

Two universes, $U_\alpha$ and $U_\beta$ are distinct if they are constructed to consist in exclusive and exhaustive alternatives or if they are consistently viewed in this fashion. Thus, $U_\alpha$ and $U_\beta$ might represent the 'state sets' of two 'machines' or the alternatives in two lotteries. If they are thus distinguished, $U_\alpha$

1. This terminology is consistently employed but a confusion is possible. With equal rectitude 'universe' may mean a 'universe of discourse'; i.e. a set of static objects, $s \in S$, pointed out in an observers (formal) language $L$. To reconcile these ideas consider $m$ tests an observer can perform on the objects in $S$. These are property tests (for properties $E, F \ldots$ described in $L$). Elementary test statements are such basic utterances as '$s$ has property $F$' (equivalently '$s$ belongs to a subset $F \subset S$') or '$s$ does not have property $F$' (equivalently '$s$ belongs to the complement of a set $F \subset S$') and the strongest $L$ statements are conjunctions like

$$(s \text{ has } E) \quad \textit{and} \quad (s \text{ does not have } F) \quad \textit{and} \quad \ldots \text{so on}$$

extended over all $m$ properties. These *strongest* $L$ statements are exclusive (only *one* can be true and *one must* be at any instant). They, or any other alternative sets formulated in $L$, describe the occurrences or options, $u$, that belong to the present universe $U$.

This $U_\alpha$ and $U_\beta$ may refer to different sets of objects $(S_1, S_2)$, to distinct descriptions of sets that have objects in common, or even to distinct descriptions of the same set. Fundamentally, $U$ is formulated in terms of tests that are made (of $E$'s value, of $F$'s value) or events that take place. $U_\alpha$ and $U_\beta$ are formulated in terms of supposedly independent tests or events.

and $U_\beta$ may coexist as simultaneously observable, being indexed (as $\alpha$ and $\beta$) by a sampling operation.

Using the notation $U(t)$ to represent a universe at time $t$ (with $t$ = time measured in observation intervals 1, 2, and so on) it is possible to establish the special equivalences ($\equiv$) of $U_\alpha(t) \equiv U_\alpha(t+1)$ . . . or $U_\beta(t) \equiv U_\beta(t+1)$ . . . and thus to count event occurrences in 'the same' universe $U_\alpha$ and 'the different' universe $U_\beta$. But the usual rider (in an alternative set) that 'one and only one of the possible events occurs *at once*' provides an essential clue.

Time is another way of indexing universes; in general, $U(t) \neq U(t+1)$. So, for example, it is quite possible, and in fact quite usual, to set $U(t) = U_\alpha$ and $U(t+1) = U_\beta$. Conversely, we may set the (Cartesian product) $U_\alpha(t) \times U_\beta(t) = U(t)$ and $U_\alpha(t+\Delta t) \times U_\beta(t+\Delta t) = U(t+\Delta t)$ (with $\Delta t$ enough time for event observation) so as to observe or talk about some configurations of (joint) events in $U(t)$ in relation to an independently sampled configuration of (joint) events in $U(t+\Delta t)$. This point is quite crucial in establishing the notion of a *frequency* of event occurrence: that is, a count of independently sampled events in *one* universe which, for this purpose, is regarded as a series of (equivalent) universes such as $U_\alpha$ defined as $U_\alpha(t) \equiv U_\alpha(t+1)$ . . . and so on.

To each alternative in a universe must be assigned a positive fractional number $p$ (say $p_i$ to the $i$th alternative) with the property that the $p_i$ summed over all the values of the index $i$, equals unity. The interpretation of the $p_i$, generally called 'probabilities', determines amongst other things who or what experiences the uncertainty or receives the information. It does not bear upon the manipulative calculus.

## 1 Multivariate Selective Information Indices

Suppose there are two or more universes $U_\alpha$, $U_\beta$, which *may* be independent. In other words, it is possible to entertain an hypothesis to this effect and to *disconfirm* it or affirm it tentatively. If so, it is possible to compute uncertainties/informations $H(\alpha)$ and $H(\beta)$. Further, by examining joint contingencies (alternatives indexed $x$ in $\alpha$ and indexed $y$ in $\beta$) it is possible to compute a joint uncertainty $H(\alpha,\beta)$. This is shown in equation 2 in appendix A. The quantity

$$T(\alpha,\beta) = H(\alpha) + H(\beta) - H(\alpha,\beta)$$

is the *transmission*, or coupling, between contingencies in these supposedly distinct, and surely distinguishable, universes. If $T(\alpha,\beta) = 0$, the universes are (tentatively) regarded as *independent*: if $T(\alpha,\beta) > 0$, they are *dependent*.

For example, the alternatives $x$ in $\alpha$ may be inputs to a 'Black Box' in Ashby's sense ($\alpha$ is a universe of inputs or a set of *input states*) while the





alternatives $y$ in $\beta$ may be outputs ($\beta$ is a universe of outputs or a set of output states). As another example, $\alpha$ is any set of alternative messages labelled by values of $x$ proper to one man or machine, the transmitter perhaps, and $\beta$ is any distinct set of alternative messages labelled by values of $y$ proper to another man or machine, the receiver. Moreover, the universes may represent the same (or different) systems at different times ($U_\alpha = U(t)$, $U_\beta = U(t+1)$) when the transmission or coupling measured by $T(\alpha,\beta)$ is, in a special sense developed in Ashby (1970), the amount of 'memorisation work' needed to establish this much coupling. In a final example, again due to Ashby (1970), $\alpha$ is the universe of prescriptive actions available to a designer who can select (in designing a plant or process) from a family, $F$, of functions, $f_z \in F$, specifying operations to be performed and $\beta$ is the universe of 'goal directed behaviours', $y$, of the plant or process as governed by its design. The example is especially interesting insofar as the degree of control which can be exerted (by the designer in this example) is known to be limited by the informational coupling. Conant's theorem (Conant and Ashby, 1970). The case cited above, is usefully compared with the case in which a designer can only operate upon the external input and output of the plant or process (i.e. upon its behaviour *per se*). The saving due to design of the first kind is gigantic.

For more than two universes it is possible to compute further coupling terms; residuals or interactions, $Q$ (equation 3 in appendix A), and thus to derive a general caculus of multivariate information analysis (the deterministic analogue for which is constraint analysis (Ashby (1964a)). The canonical inequalities of multivariate information analysis are stated as tersely as possible in Ashby (1969).

One extremely useful index is the redundancy, $Z$, of a given universe (equation 4 in appendix A) which is a measure of constraint, either deterministic or probabilistic; *i.e.* the extent to which the universe is coherent or organised. As noted in the Appendix, there are many plausible definitions of redundancy, but all of them serve as indices of uncertainty and are functions of at least two variables. One indexes the amount of variation that might occur within the stipulated constraints (if any) upon the universe, the other indexes the variation that is manifest. In other words, the redundancy is a measure of the transmission between a specification of *one* universe and the events occurring within it.

If it happens that the $p$ numbers can change over time (for instance, the 'time and memorisation' example given above) and it also happens that the *universe* can change over time (either because *it* or its irreducible constraints are modified) then there may be a non-trivial form of self organisation (equation 5 in appendix A).

Call any or all of the quantities $H$, $T$, $Q$, $Z$ or others derived in a similar manner 'uncertainty/information indices'. In general, they are *selective* information indices since information is gained by selecting amongst a set of alternatives. The interpretation of these indices depends upon the genesis of the alternative sets in the universe(s) concerned and upon the origin of the $p$ numbers.

## 2 Statistical Indices

An external (impartial, unbiassed) observer may overlook a system which in the last resort *he* has defined; for example, the input set and output set of an organism; stimuli it receives and responses it emits. Further, he may compute the $p$ numbers in any systematic manner. He could, for example, estimate them by guessing (if he were a Bayesian statistician he could do so rationally) or, more often, he subscribes to the common statistical dictum that a probability $p_i$ (that $x$ has the value $x_i$) is obtained, as the limit, for infinite sequence length of the ratio

$$\frac{\text{Number of occurrences of } x = x_i}{\text{Number of occurrences of any value of } x}$$

The formulation depends upon 'stationarity' of the source of occurrences (a notion considered further in Chapter 2). Roughly, 'stationarity' means that the underlying structure of the source is unchanging.

Under this interpretation, the uncertainty/information index is a *statistical* uncertainty/information measure. Given either method of computing the $p$ numbers, any uncertainty information indices reflect *his* uncertainty/information about the system. For example, subscribing to the last dictum, the external observer may record event occurrences in a contingency table to show the frequency with which events tagged by values of $x$ (input states) and $y$ (output states) occur, including the events due to joint occurrences. The entries in the contingency table may now be taken as estimates of the actual probabilities (a good example is given in McGill (1963) who also provides the mathematical apparatus needed to compute in terms of count numbers without transforming them into $p$ numbers). Some instances of this type of calculation are cited in Chapter 7.

Next, an external observer may presuppose that some other observer (a participant, like an operative in industry, or an experimental subject) sees an *environment* in the same way that *he* sees it. If so, he may infer that the *statistical* uncertainty/information indices computed with reference to *this* environment provide estimates of its variability, complexity, regularity, or whatever (depending upon the chosen index), from the participant's point of view. Some examples of large-organisation task-analysis along these lines are shown in Table 1 and Fig. 1. Similar studies have been carried out





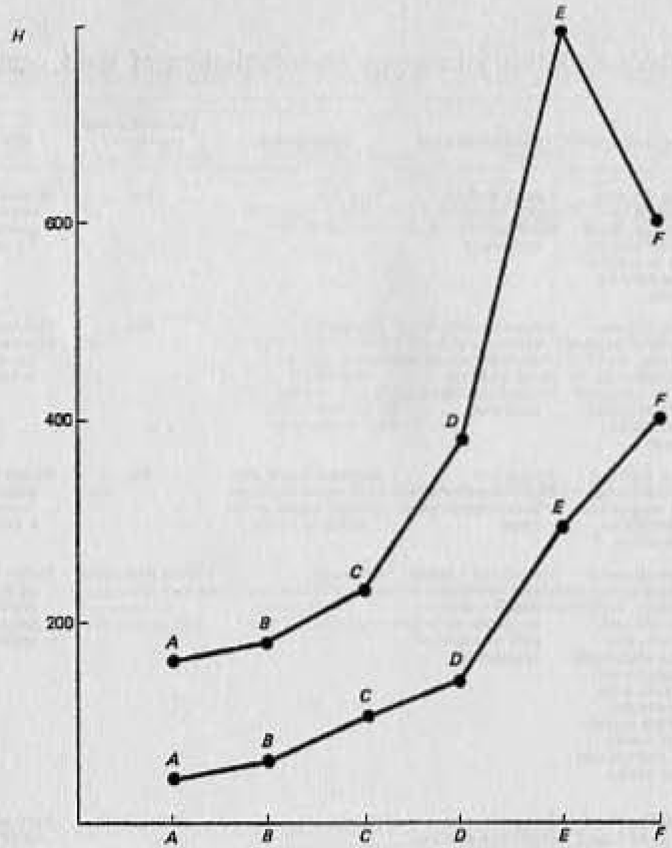



*Figure 1* Information analysis for six industrial/commercial tasks *A, B, C, D, E, F*. Upper curve (appropriate to novice) is gross uncertainty. Lower curve (appropriate to a skilled operative) is uncertainty, given that all constraints are known. Difference (a transmission) is an index of the amount to be gained by experience. The lower curve is an index of relative difficulty of tasks for any (idealised) proficient operative.

in the timber industry by Van Gich (1971) who relates the indices to ergonomic and work study measurements. Stanisland (1966) has used similar techniques in connection with laboratory tasks and has devised a plethora of redundancy measures for this purpose.

## 3 Selective Operations and Selective Work

Suppose that $\alpha$ and $\beta$ are the input and output sets of an organism, perhaps a human being (as an expository convenience; the argument is quite

*Table 1* Ordinary language characterisation of tasks compared in terms of information indices in Fig. 1 (study by: Scott, Elstobff and Pask).

| *Job label* | *Brief description* | *Outline structure* | *Interruptions* | *Communication important?* | *Effect of errors* | *Repetition* | *Decision making* | *Input to the process* | *Special knowledge required* | *Specialised vocabulary used* |
|---|---|---|---|---|---|---|---|---|---|---|
| *A* | Checking the correctness of invoices, etc., from suppliers and invoices as correct for payment by treasurer | 3-stage process<br>Single operator<br>Batch processing of each stage | Very few | No | Moderately serious since operative is passing invoices for payment | Highly repetitive | No | All input available at start of process | Little special knowledge required | Small specialised vocabulary used<br>About 15 terms |
| *B* | Collation and preparation of batches of invoices for processing prior to input to computer<br>Checking of copied invoice against originals | 3-stage process, each operative able to carry out all stages<br>Batch working<br>Parallel working at each stage | Infrequent | No | Not very serious<br>Most errors picked up at later stage in processing | Highly repetitive routine work | No | Input to the process is complete at the start of the process<br>No information gathering or waiting for information required | Very little special knowledge required | Quite small specialised vocabulary<br>About 25 terms |
| *C* | Updating staff list with information about resignations, new employees and transfers | 2-stage job<br>Single operator batch processing of each stage | Frequent (clerk also operates telephone switchboard whilst doing this job) | No | Errors important since this job involves updating a master record | Highly repetitive | Some | All input available at start of process | Little special knowledge required | Small specialised vocabulary used<br>About 10 terms |
| *D* | Determining customer's requirements for service, provision removal, cessation, etc.<br>Checking equipment and plant available, arranging appointments, obtaining acceptance of conditions, charges and issuing advice notes | Average of 5 stages<br>Operators work in parallel, each responsible for a part or whole of organisation | Frequent | Very important | Errors fairly serious as work of other divisions depends on operator's output | Varies as to areas served<br>Many cases similar, others unusual or infrequent | Difficulties may be referred to superintendent | Information gathering is necessary<br>Waiting often required | Considerable special knowledge required | Large specialised vocabulary used<br>About 200 terms |
| *E* | Preparation of legally required load documents for a craft, using information on load plan | 4-stage process<br>Operators working in parallel each operator responsible for one flight at a time | Occasional | Occasionally | Very serious: safety laws demand accuracy | Repetitive process largely routine | No | No information gathering necessary, but waiting for required information is necessary | Considerable special knowledge required | Large specialised vocabulary used<br>About 150 terms |
| | Assembly of data and preparation of a loading plan | 6-stage process<br>A separate operator carrying out each stage<br>Sequential processing, i.e. each stage being fed by its predecessor<br>Concurrent processing of load plans within some stages | Frequent | Yes | Very serious: safety laws demand accuracy | Repetitive process but not routine<br>The load planning of each flight is a separate problem | Yes | Information gathering necessary<br>Waiting for required information necessary | Considerable special knowledge required | Large specialised vocabulary used<br>About 150 terms |

general). Let $x$ (indexing elements of $\alpha$) be determined by an external agency and consider the value of $H(\beta)$, the output uncertainty. If there are just $M$ values of $y$ (indexing the $M$ elements of $\beta$) then surely $\log M \geq H(\beta)$ with equality if the alternatives are equiprobable. Now envisage the organism carrying out a selective operation in the following sense; it is designed, or has evolved, or is told that, for each value of $x$, there is a correct value $y = R(x)$ of $y$ and it has the 'goal' of selecting this value of $y$ for each $x$. The selective operation it performs entails so much selective work and if we hypothesise that a general dichotomising process takes place which occupies a unit interval, then the larger $M$ is the larger will be the interval (the reaction time or the latency) for any correct response.

Hick's (1952) law asserted such a correspondence which was later established by Crossman (1953) and others; that is, if the correct reaction time or latency is written $RT$ then

$$RT = \text{constant} \times \log M$$

In fact, by using frequency-weighted selection categories, it is possible, for some tasks, to assert that

$$RT = \text{constant} \times H(\beta)$$

or (as appropriate, on varying $R$)

$$RT = \text{constant} \times T(\alpha,\beta)$$

though the results (obtained, for example, by Fitts (1954); Bricker (1955)) are sometimes enigmatic. The deviations from Hick's law are primarily due to the fact that (apart from the type of experiment Hick employed in his empirical studies) the external observer's computation does not tally with that of the subject or organism. However, it is still true that when $H(\beta)$ is less than $M$ because the selection probabilities differ, the correct selection can be expressed as a selection from a number, $EM$, of (imaginary) equiprobable alternatives equivalent to the actual set of biassed alternatives, and the selective work is measured by $EM$; that is

$$RT = \text{constant} \times \log EM = \text{selective work}$$

The notion is not trivial. The selective operation is a *non-specific computation*. It (and information theory as a rule) makes no comment upon *how* a computation is performed; given $x$ there are indefinitely many ways of computing the required value of $y$. The index of selective work is an average taken without discrimination over all of these possibilities, and selective work, indexed by $RT$, may be taken as a gross index of the uncertainty that





must be resolved by a subject regardless of whether or not the external observer's view of the environment is in agreement with the subject's view.

Similar, but differently phrased, arguments underlie the excellent work of Tanner and Swets (1954) on signal detection in a 'noise' background.

The selective work index is used implicitly in expressing latency-weighted measures of task performance in terms of information measures, in Chapter 7 and Chapter 8 especially.

## 4 Capacity Limits

One of the main results in the commonest application of statistical and selective information theory (namely, systems composed of a receiver and transmitter coupled by a channel which is imaged in the abstract by a transmission) is the limiting channel capacity theorem (Shannon and Weaver, 1949). It is assumed that perfect reception of a message is hampered by the introduction of 'noise' irrelevant to the transmission; it is also assumed that the receiver and transmitter have statistically stationary properties so that conditional and joint probabilities can be estimated by ratios formed on averaging over long sequences of inputs and/or outputs.

The capacity theorem says that it is possible to encode messages, transmitted along the channel (i.e. to impose a structure upon the transmission which relates values or sequences of values of one variable ($x$) to the values that are assumed by another variable ($y$)) in order to secure nearly error free transmission as the rate of transmission is increased (Fig. 2); beyond that limit (the 'capacity' of Fig. 2) distortion is increased by any increment in rate.

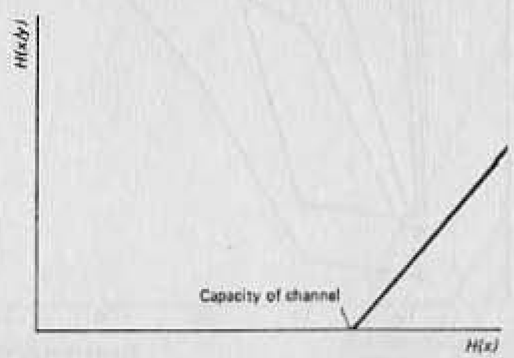



*Figure 2* By suitable encoding the channel may transmit without distortion (or error) for any rate up to the capacity but, beyond that, the minimal error increases with increasing rate. $H(x/y)$: uncertainty about value of input $x$ if value output $y$ is given, $H(x)$: input information rate or uncertainty about value of $x$.

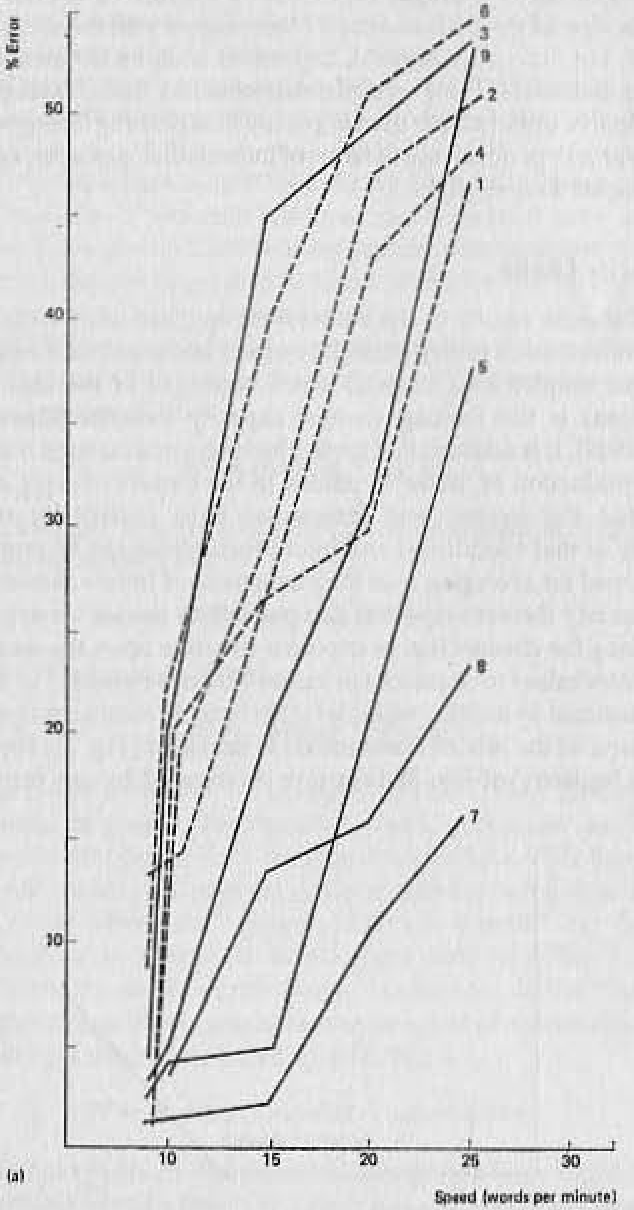


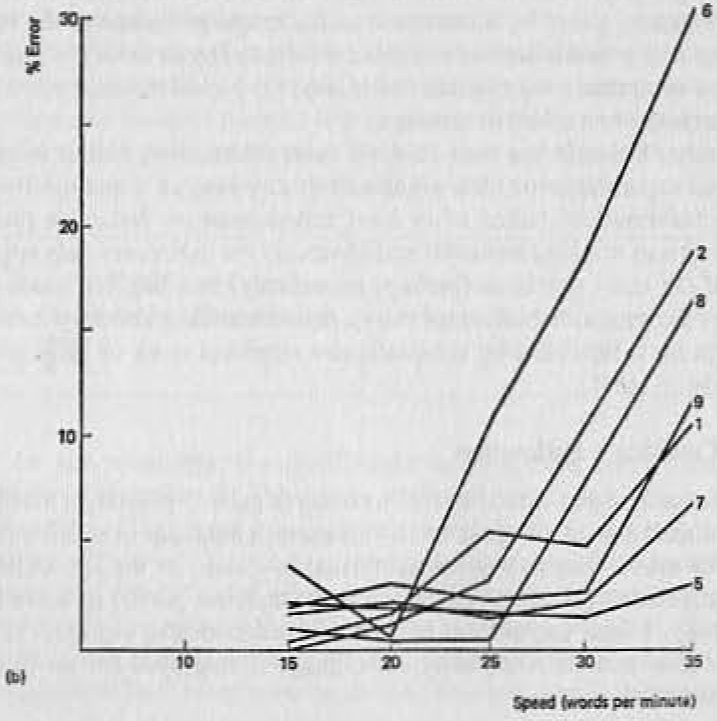

*Figure 3* Pace error curves: (*a*) at an early stage of training; (*b*) at a later stage of training.





## 5 Human Capacity Limits

Some human experiments or task situations can be associated with plausible criteria of relevance and it is also approximately true that the human being has stationary characteristics (that is, a fixed processing structure underlying statistical variations). This is true, for example, of many psychophysical experiments where the human being is regarded as an essentially physiological device which can be imaged as a noisy channel that couples an input ($x$) to an output ($y$). When the human being learns and when he performs symbolic transformations it is neither possible to specify relevance (in an adequate sense) nor to assume stationarity. Hence, under these circumstances a human being does not have a channel capacity in the strict sense. But it is often true that Hick's principle (in precis, the 'rate of gain of information is constant') determines a *loading* which is analagous to a channel capacity. The loading limit is particularly evident in the context of paced-input skills, like teleprinting in class with key-depressions synchronised by a metronome beat. The absolute degree of selection depends upon the trainees level of competence and (generally) increases as he learns. But, for all trainees who do learn, there is a marked inflexion in the error/pace



graph. Beyond a certain fixed pace, any trainee commits errors. If he is able to compensate for errors up to a limit at which breakdown occurs (by analogy only, by effective 'encoding'), there is a discontinuity as shown in Fig. 3a (by analogy only, the 'capacity'). In contrast, some trainees commit errors at any pace of operation and there is no real limit. These trainees also fail to learn the repetitive skill (Fig. 3b shows error pace curves for the same subjects, after a further 18 hours of training).

## 6 Logical Uncertainty/Information Indices

The general (method-of-computation independent) index 'selective work' is usefully contrasted with an information index which is, in one sense, its polar opposite—an information measure based upon one class of possible computing methods. Measures of this kind are called logical information measures and are specified in respect of a formal language which is very restricted; the class of computations being the class of derivations obtainable by using the inference and production rules in this language. These measures were devised by Carnap (1950), Carnap and Bar-Hillel (1953) and Bar-Hillel (1964). Some important generalisations are due to Chiaravigillio (1971), and Harrah (1966). The strongest statements in such a language, together with their complements, form alternative sets (the 'information given' by a statement in the language measures the number of logically possible derivations that are excluded by virtue of the statement) but, *if* computed over alternative sets *only*, the logical measure corresponds numerically to a selective measure.

Rather little use has been made of these information indices in psychological experiments or task situations, at any rate at a quantitative level (the 'information' talked of in most experiments on deductive problem-solving is of this *kind* however) and obviously the indices are only appropriate if the mind functions (perhaps imperfectly) in a 'logical' mode (*i.e.* it obeys the prescribed inference rules). An outstanding and very interesting exception is furnished by the *qualitative* empirical work of Seigmann and Stapleton (1971).

## 7 Confidence Estimation

Consider an agent who is a sentient being (a man or possibly a machine or an animal) able to interpret and comprehend a language in which questions can be asked. The very restricted formal languages of the last section can be augmented (as in the Seigmann and Stapleton paper) to serve in this capacity. I shall use natural language in order to give examples; in fact, some intermediate complexity of language is employed for much of our current work.





The agent is asked a multiple-choice type of question. Syntactically, it is an exclusive disjunction of statements. But these statements have a semantic interpretation whereby the agent refers them to a body of internal or external data and the question is prefaced, when it is addressed, by the *requirement to satisfy* a relation (which is one meaning of the many-faceted and often misused word 'goal'). Succinctly, a question of this type poses a problem; to find (say) the one and only one alternative answer, satisfying the relation 'bridge-building engineer', within an interval, $\Delta t$ (which may be indefinite), given a semantic interpretation of the language in respect of industrial history and the following alternative statements:

(a) 'Brunel built the Bristol Suspension Bridge',
(b) 'Stephenson built the Bristol Suspension Bridge',
(c) 'Wakefield built the Bristol Suspension Bridge'.

The semantic interpretation admits, for example, such factual data as: 'the Bristol (Road) suspension bridge was built by an engineer who built railways' and 'Wakefield built no railways' and 'Stephenson did not design Temple Meads Station at Bristol'; such inferences as 'Wakefield did not build the bridge' and 'if it is true that the bridge was built by an engineer who built a railway to Bristol then it is not false that Brunel built the suspension bridge'. The pragmatic interpretation has an imperative/interrogative part (a question of this kind is isomorphic with a command to answer) and it specifies the relation deemed 'correct' (namely between *road* bridges and *builder's names*). It is improper to gloss the issue by conceiving the question as the stimulus for a correct response, even though this identification is suggested by the usual *format*, namely:

| | |
|---|---|
| Who built the Bristol Suspension Bridge: (a), (b), or (c)? (one and only one is right) | (a) Brunel<br>(b) Stephenson<br>(c) Wakefield |

**7.1** As one possibility, the agent might *derive* a statement to resolve the exclusive disjunction (a), (b), (c), by computation in the language (the object language of the scheme), using inferences of the kind suggested above and starting off from the existing and comprehended data. As a comment, if he did so and merely ticked or marked the alternative corresponding to his derived (alias computed) object language statement, then this kind of enquiry gives little of the potentially available information. The agent might, for example, have been asked how he arrived at his solution to the problem.

However, this possibility depends upon the existence (in the agent) of a computing method and upon there being enough time to execute it; that is, $\Delta t \geqslant$ object language computing time.

**7.2** If the agent had *no* method, then he could not compute an answer. Since one answer must be given he may either construct a method or guess an answer within $\Delta t$. The former activity is learning and the latter is a minimal instance of choice or decision.

**7.3** If a computing method is available to the agent, without the necessity of constructing one by learning, then its application still depends upon a sufficiently lengthy $\Delta t$ (*how* long is determined by the method, the agent's state of knowledge and so on). But even if $\Delta t$ is too short, the agent may carry out a partial computation that eliminates some of the alternative statements; as a result he may be able to guess between fewer alternatives and thus to make a more discriminating choice. (Parenthetically, 'cueing information', delivered in the course of computation, is information in the logical sense; cueing information that excludes the *names* (a,b,c,) of the alternatives is selective.)

**7.4** Suppose there is a *metalanguage* (able to accommodate descriptions of classes of object language statements, inference rules, etc.). In particular, suppose the metalanguage is a statistical metalanguage (formal or not) of which the formal grammar consists in rules of statistical inference; for example Baye's rule. (This example is due to Phillips, 1970.)

If so, the agent is in a position to compute weights or likelihoods attached to the alternatives, and to bias any otherwise random guess (in the limit by giving one or more alternatives a weight of zero but, in any case, refining his choice or decision).

**7.5** The data treated 'statistically' (either by formal precept or rules that are not fully axiomatised) may be derived from partial computations, relevant to the questioned relation, that are carried out in the object language.

**7.6** The data might, however, be quite different; for example, the frequency with which alternative statements that happen to be labelled '*a*' have been deemed correct at previous trials.

**7.7** There is also a special type of distortion, noted by Shuford (1965), Shuford, Massengill, and Albert (1966) and Finetti (1962) which is obtrusive if correctness is assigned a score value, if the answers to a test sequence of questions are scored cumulatively by correctness, and if the agent aims to





maximise his score. Acting as a good statistician (at any rate) he will appreciate that the mathematical expectation of score is maximised by a deterministic choice of the most likely alternative and not by a biassed guess between alternatives weighted to represent a current state of knowledge; from an observer's point of view this clever calculation is *irrelevant* to the agent's state of knowledge. (See Appendix B.)

**7.8** With 7.5, 7.6 and 7.7 in mind, it is possible to ask the agent for a confidence estimate over the alternatives instead of a selection. This is the result of a metalinguistic computation, constrained by the metalinguistic syntax to yield as many positive or zero fractional numbers as there are alternatives, with the property that all of them sum to unity, i.e. these numbers have the form of probability numbers, $p_i$, attached to the members of an alternative set (in the example $i = a$ or $b$ or $c$). In other words, the confidence estimate is the output of a process that would, if selection were demanded, bias the otherwise random throw of a 'mental dice' with as many sides as there are alternative statements

**7.9** Under certain circumstances, confidence estimation is impossible or uninformative. Presumably a metalinguistic computation occupies a certain interval of time; thus the estimate cannot be elicited unless $\Delta t \geqslant$ metalinguistic computation time. Again, if the object language computation time is shorter then the metalinguistic computation time (and an object language computation or derivation method exists) the appearance of confidence equal to unity in one alternative (by the imposed syntax, the other $p_i$ being zero) is due, in fact, to the object language derivation of a statement. If $\Delta t$ is too short either for object language computation or metalinguistic computation, then the agent is bound to guess in any case. But, in general, the confidence estimate gives more information to the experimenter or external observer than a direct selection and, insofar as the pathologies noted in sections 7.6 and 7.7 are avoided, this is relevant information.

**7.10** The confidence estimate *itself* is interpretable as the agent's 'degree of belief' in each of the alternatives and is usually elicited by a request to express degrees of belief in terms of a betting or wagering distribution. Under this interpretation an uncertainty information index, calculated over the alternative set, using the degrees of belief $p_i$, is an index of subjective or 'to the agent' or 'as seen by the agent' uncertainty/information. All of the uncertainty/information indices used in Chapter 11, section 3 onwards, are subjective indices.

The argument of this section is summarised in Fig. 4.

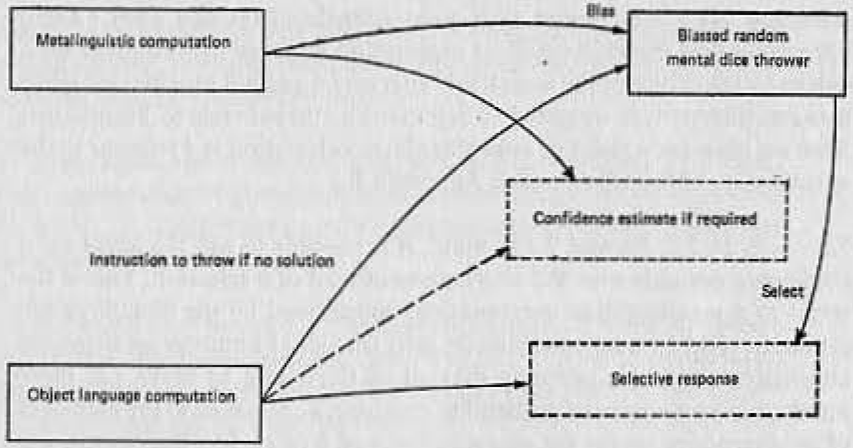



*Figure 4* Summary of discussion of confidence estimation (see text). Dotted line represents exclusion of alternative statements by deletion buttons on BOSS.

## 8 Methods

Various precautions are taken to ensure that data of this kind, indicating subjective uncertainties/informations, is not spurious.

**8.1** Concentrating upon the pathology of section 7.7, Shuford and his colleagues have devised schemes that defeat the 'good statistician' trick. Two are outlined in appendix B (for the minimal case of two alternative statements).

**8.2** Clearly, an agent can be certain and correct or certain and mistaken; in doubt and favouring the correct alternative or in doubt and favouring a mistaken alternative. Only if he assigns equal probability to each alternative is the issue of correctness immaterial.

**8.3** Hence, indices of subjective uncertainty/belief are augmented by an index of correct belief calculated from the same data. A Shuford score (the simplest of them all) is used for this purpose:

$$\text{correct belief} = \begin{cases} 1 + \log p_i \text{ for } p_i \geq 0{\cdot}1 \\ -1 \qquad \text{for } 0{\cdot}1 < p_i \end{cases} \left(\begin{array}{l}\text{if the } \textit{correct} \\ \text{alternative is} \\ \text{the } i\text{th}\end{array}\right)$$

**8.4** It is possible to guard against irrelevant estimates of section 7.7 in several ways; one of them is the Shuford method since, in fact, the statistician's maximising strategy is a valid but irrelevant metalinguistic computation and the act of thwarting it discourages irrelevant computation in general. Other techniques are noted below.





**8.5** The burden of satisfying the metalinguistic 'syntax' (that any well formed statement for $M$ alternatives is a set of $M$ numbers $p_i$ such that $1 \geq p_i \geq 0$ and $\sum p_i = 1$) may be appreciable. To relieve it, the syntactic constraints are satisfied externally by a normalising equipment which only permits 'well-formed' statements. Baker (1969) uses a histogram; one bar to each alternative, bar heights representing degrees of belief. It is displayed on a computer controlled cathode-ray tube. The heights of the histogram bars are adjusted at will, one at once, by the agent; but the bar heights are automatically normalised so that they sum to unity. A very simple equipment will suffice for many purposes. The $p_i$ are displayed on meters wired in series with variable resistances and fed from a constant current source; the agent adjusts variable resistances (one to each meter) to change his asserted degree of belief.

**8.6** The crucial requirement is to reduce the metalinguistic processing time since (from section 7.9) this favours the various sampling conditions that are propitious from the external observer's point of view.

Provision of an external normalising circuit is a step in the right direction. Some improvement is effected if the equipment is automatically reset for each questioning trial. It is still better to complete the externalisation by providing a feedback signal indicating the degree of freedom used up (or the constraint imposed by the agents previous adjustments) in order to arrive at his confidence estimate.

Since the external observer does not, and cannot in this format, discriminate between *relevant* metalinguistic computations and object language computations, it is important to allow for the rapid exclusion of alternatives as a result of object language computations that are partially carried out. All of these specific expedients also act as devices that encourage the agent to produce $p_i$ values by relevant computation. They do not ensure that he will do so; that depends upon the normative scheme in which the agent accepts the goal of treating questions as problems that are solved by satisfying a goal relation.

**8.7** The sampling equipment (Belief and Opinion Sampling System, BOSS) used at present in my own laboratory is shown in Plate 1. Questions are inscribed on cards bearing code holes for various data, including a designated correct alternative. Insertion of the card resets the circuitry and activates those out of the alternatives (maximum of 8) that are in use at a given trial. Degrees of belief are displayed on the meters which are initially assigned equal readings. The agent adjusts this profile of readings by pushing levers to increase or decrease meter readings corresponding to each of the active alternatives. Members of the active set may be excluded by pressing deletion buttons. The profile is automatically normalised and the 'existing constraint' feedback is simulated by a variable lag or delay in the meter response. When the agent is satisfied with the profile he submits it for evaluation by the sampling system; if appropriate, he receives knowledge of results, and is permitted to withdraw the question card and insert another one.

**8.8** Two further points should be stressed.

(a) The confidence estimation paradigm does not demand an external criterion of correctness; it is only necessary that one alternative be selected. So, for example, the agent may be asked to express his belief that he will purchase one and only one of several alternative products, in which case the only correctness criterion is internal, i.e. his personal preference. To distinguish trials for which there is *no* external correctness criterion, they are called option trials and the profile of readings is referred to as an 'opinion' statement rather than a statement of *belief*.

(b) The paradigm does not demand that the alternatives (provided they *are* exclusive and exhaustive) have been chosen by an experimenter or an external observer. Unless inter-agent comparisons are essential it is convenient to allow the agent to select his own sets of alternatives quite freely (he inscribes them on blank question cards and inserts them himself). Often, also, it is convenient actively to solicit the agent to specify his own alternative statements (for example, in studies of strategic uncertainty, noted in Chapter 11) and several variants on the repertory grid technique (Kelly, 1955) have been used for this purpose.

## 9 Experimental Examples

The following instances illustrate the use of uncertainty/information indices based on confidence estimation (interpreted as indices of subjective uncertainty/information).

**9.1** Figure 5 shows an information analysis of a clerical operation. Given an invoice (and invoices are processed in lengthy batches) the clerk must find and record the code numbers of the several items referred to. The item codes are specified in a code book in which items are grouped under categories and, from this book, it is possible to compute the actual dependency between the relevant variables (supplier's name, origin name, area, item group, item, and so on). However, it would be impracticably tedious





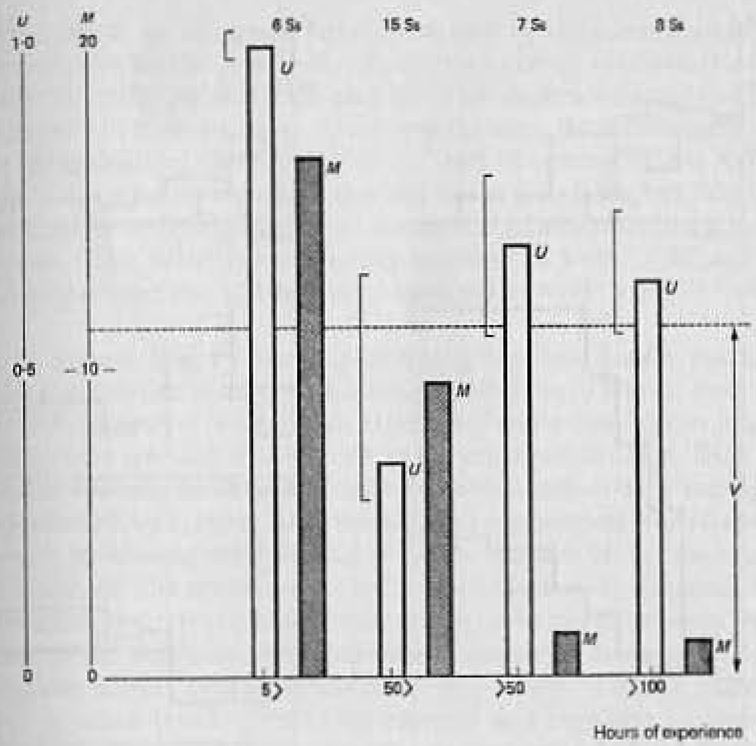



*Figure 5* Information analysis of a clerical operation: $U$ is average uncertainty, given a document regarding the values of code variables. The uncertainty remaining, if the contents in a code book were known perfectly, is designated $V$. Values of $U < V$ indicate spurious overlearning. Points are plotted for novices, clerks with about fifty hours of continuous experience, usually obtained in hourly spells over a month or more, and fully experienced clerks with more than a year's experience. $M$ is a gross estimate of the number of mistakes or omissions at each level of experience.

for anyone but a novice to engage in a 'look up' operation except for a rare or ambiguous item, and, as they become more proficient, trainee clerks do not do so. The uncertainties they experienced, at various stages in training, about the values of the variables when initial values are given (on picking up the input document), were determined by confidence estimation over ranges of possible values. The actual uncertainty is an average over stages in a single pass of the coding operation since variable values are determined in stages and fresh transmission terms have to be considered. In practice, it is impossible to obtain confidence estimates for all trials (the clerk would have no chance to do his work!) so confidence estimates were sampled occasionally, on marked invoices.

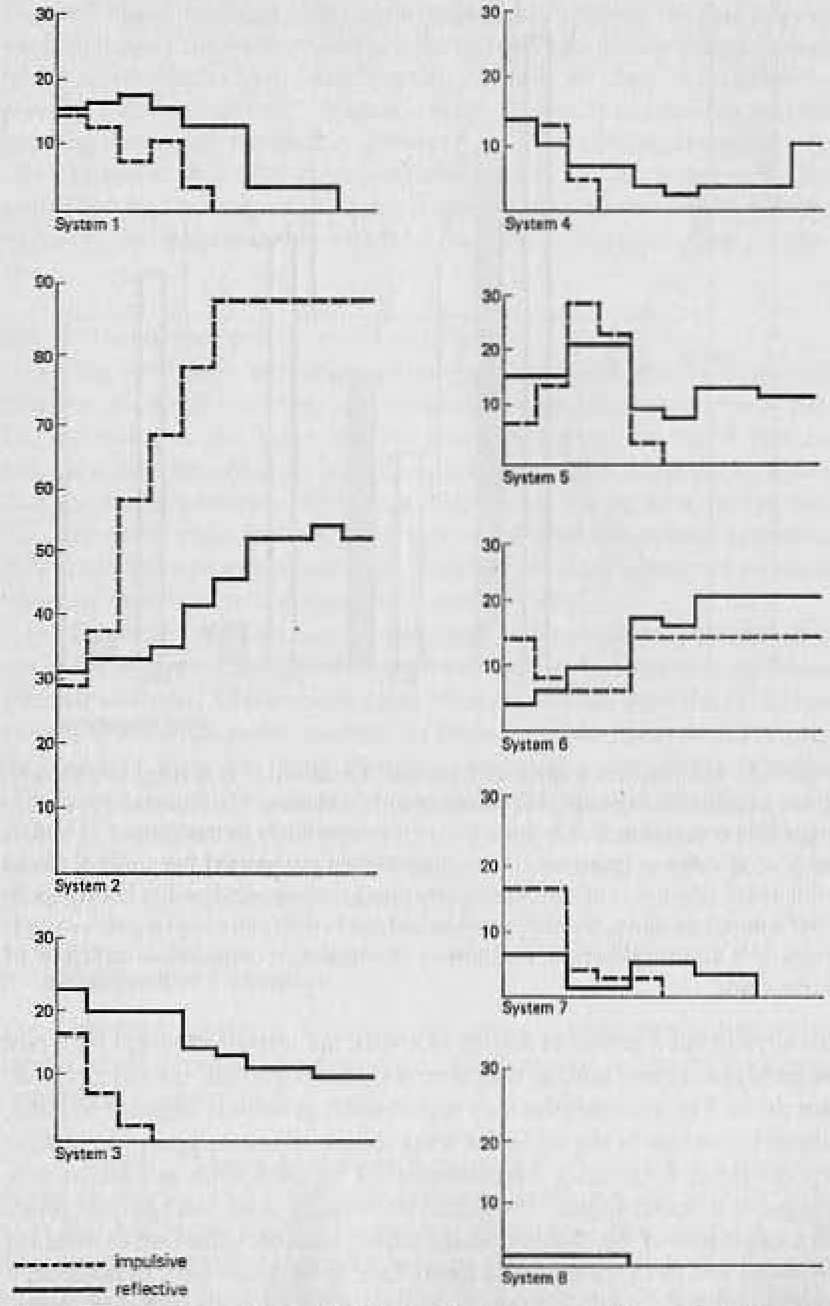



*Figure 6* Time changes in probability (of choosing one or other system) averaged over impulsive and reflective groups. Abscissae, time; ordinates, subjective probability of purchasing the system. - - -, impulsives; ——, reflectives.





Each bar in the histogram represents a level of experience (number of hours of prior familiarity with the job) and is an average over data from the number of clerks (Ss) cited at the head of it. The brackets indicate 1 standard deviation unit from the mean. Apart from learning, the dominant effect is over-generalisation. Clerks at the '50 >' level of experience have learned many of the genuine constraints specified by the code book, but they have also imagined or falsely inferred the existence of general constraints that do not exist. This defect is substantially remedied at level '> 50' and the necessary discriminating details are established by level '> 100'.

**9.2** Figures 6 and 7 show opinion-sampling data from a purchaser decision-simulation conducted with equipment shown in Plate 2. Respondents with an interest in buying an expensive consumer durable (a heating system) were recruited and brought to the experimental room. Each respondent was independently classified by tests for a 'reflective' or 'impulsive' disposition (Kagan, 1966). After briefing, they spent several hours (between 2 and 4) in choosing between types of system available on the market and one 'catch all' (the system owned by the respondent at the moment). It is, in fact, true that one purchaser would only want to buy one system. In the course of the simulation run, confidence estimates (of the 'opinion' type; there is no 'correct' system) were obtained periodically.[1] Figure 6 shows the mean $p_i$ values ($i = 1, \ldots, 8$) for reflective and impulsive respondents separately. The uncertainty/information indices are calculated from the individual data and are shown in Fig. 7. The simulation itself is an arrangement that allows and encourages the respondent to explore veridical sources of data; advertising, advisory services, comments of other users, informative brochures, graphic material and so on. It is also possible to institute purchase and installation arrangements to which the proper delays are attached. The simulation is paced by a 'clock' which runs at a rate determined by the current transaction (exploratory or otherwise) but, on average, condenses a four month decision period into three hours (probably about half the interval spent, during the four months, in relevant decisive activity).

The transactions are recorded. It would be possible to estimate information *transmissions* from the data gained in exploration of the data base

1. In fact, the respondent is periodically requested to state his goal (i.e. the domestic requirements he chiefly wishes the system to satisfy) by setting rank ordering controls which determine his requirements as a region in a 'space' of product 'dimensions' (the character of which was determined by extensive field studies using interviews with respondents and a repertory grid technique). The same dimensions cross index the data base. Further, goal data, as well as confidence estimation data, are fed into an on line computer model which, in addition to record keeping, predicts the respondent's choices and provides him with recommendations. All this, however, is part of a larger story.

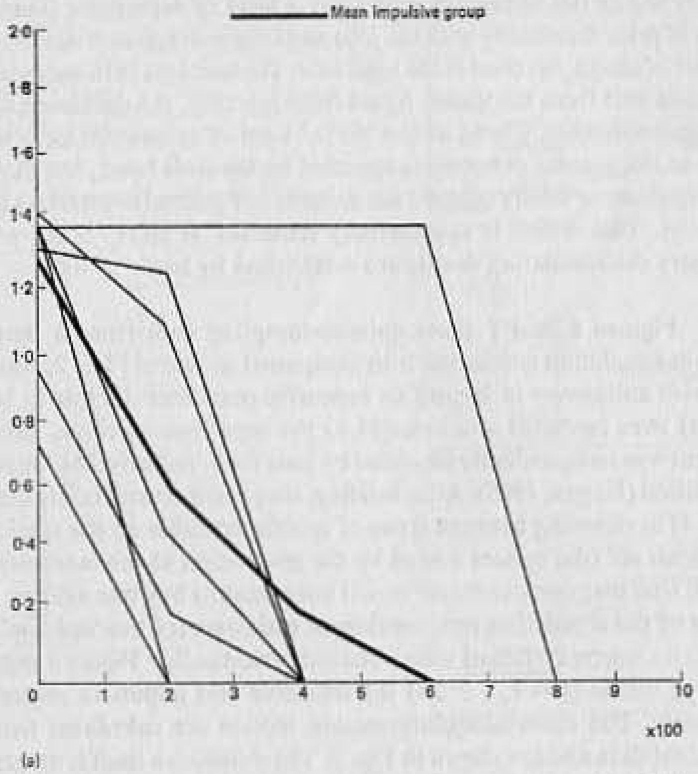


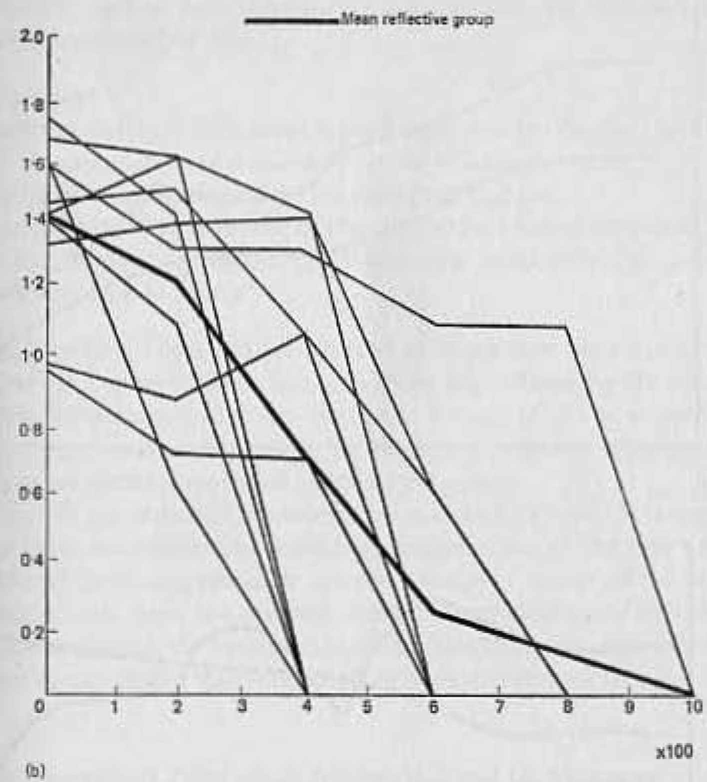

*Figure 7* Uncertainties for each respondent and mean values for: (*a*) impulsive group; (*b*) reflective group. Abscissae, time; ordinates, subjective uncertainty.

by auxiliary confidence estimates. For this simulation the transmissions were estimated by reference to the microscopic statements in each part of the data base and on the assumption that any piece of data obtained in exploration was noticed and retained.

**9.3** Figure 8 is a record of belief-sampling; that is, confidence estimation in which a correct alternative is specified for each trial. The questions refer to items in a data base (very similar to the decision simulation data base of the last section) over which people have a measure of doubt. As might be expected, the provision of knowledge of results after each trial gives rise to a regular reduction in the uncertainty index and a corresponding increase in correct belief (the index specified in section 8.3). In the absence of knowledge of results a respondent becomes increasingly familiar with the items and his uncertainty also is reduced; but his correct belief is not always in register, i.e. the respondent becomes certain about falsehoods.








**9.4** Quite complex experimental designs are feasible and may be used to tap *opinion* about such options as products to be purchased, products to be developed, or politicians to be elected; or to ascertain *belief* about the qualities of a product or the actual implications of an asserted policy, or, for that matter, any confirmable issue.

For example, in one study (forty-seven respondents in the socioeconomic $C_1/C_2$ category) it was found that *if* the effect of advertising produces a change in opinion it always leads to *increase* in the uncertainty of choice. As a practical corollary, once opinion has been changed in favour of a product (or a politician who is promoted or a given line of propaganda) then, to establish this change firmly in mind, confirmatory data must be provided. After that, due probably to cognitive dissonance in respect of the changed hypothesis (Festinger, 1957), advertising in favour of the original product has little effect upon opinion. Some data (with products obfuscated but otherwise of the correct order of magnitude and numerosity) are shown in Figures 9 and 10.

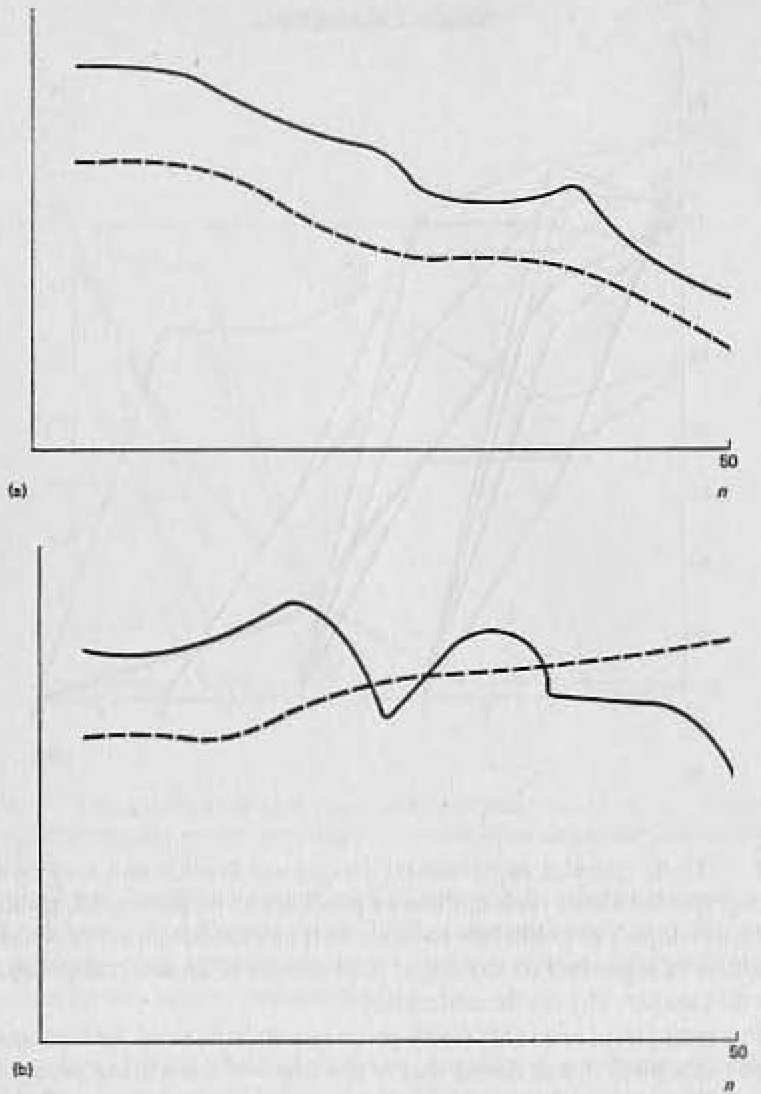



*Figure 8* Belief (as uncertainty) and correct belief as a function of repeated questioning: (*a*) with knowledge of results; (*b*) without knowledge of results. Plain line, uncertainty ($H$); dotted line, correct belief; $n$, number of trial.

The experimental session begins with pretraining in the use of BOSS (enough to secure probabilistic judgements over genuinely risky or aleatory alternative sets and also *change* of opinion due to *conditional* information). The pretraining materials are now set aside and successive judgements over





*development options* and *purchase package options* are recorded in the following sequence (see Fig. 9).

1. Initially;
2. after sampling beliefs about factual aspects of the product field;
3. after exposure to advertising in favour of a new product;
4. after an attempt to confirm the change of opinion;
5. after advertising in favour of the original (not the *new* product) and;
6. after training in respect of factual data *with* correct or mistaken 'knowledge of results'.

The last (training) operation is inserted to check that the measured 'uncertainty' is a quantity that is, as it should be, reduced by the receipt of objective information. Opinion change in favour of the new product is clearly perpetuated even though the original product is advertised (the products are alternatives for all practical purposes).

In Fig. 10 the order of presentation of materials (4) and (5) is reversed. Under these conditions (no immediate confirmation of the new product hypothesis) there is appreciable opinion change in favour of the original product. Hence, there is a reversal. Since the new product advertising has not been confirmed, the presentation of data favouring the *original* product leads to an ambiguity, that is resolved by reversion to the original.

## 10 A Theoretical Point about Information and Its Measures

It is well known that the mathematical form of expression used to represent a selective uncertainty/information index mirrors the mathematical form of an *entropy*; that is, the physical quantity which increases over time in a physical closed assembly as energy is degraded into less specific and usable forms, for example, in the operation of a heat engine or the metabolism of a cell where work is done. The correspondence comes about because, at a molecular level, the increase in entropy can be expressed in terms of an increasing disorder amongst such systems as motions of particles, distributions of elements, as they appear in the descriptive framework of physics. The quantity that decreases when entropy increases '– entropy' is called *negentropy* and as argued by Brillouin (1956, 1964) or, in the domain of living systems, by Schroedinger (1944) any *use of* information demands an increment in negentropy which must be furnished by the provision of energy or degradable materials, such as a foodstuff. Brillouin makes a distinction between 'Bound Information' concerned with the configurations of a physical system and 'Free Information' concerned with the possible configurations of an abstract system having no specially physical interpretation.

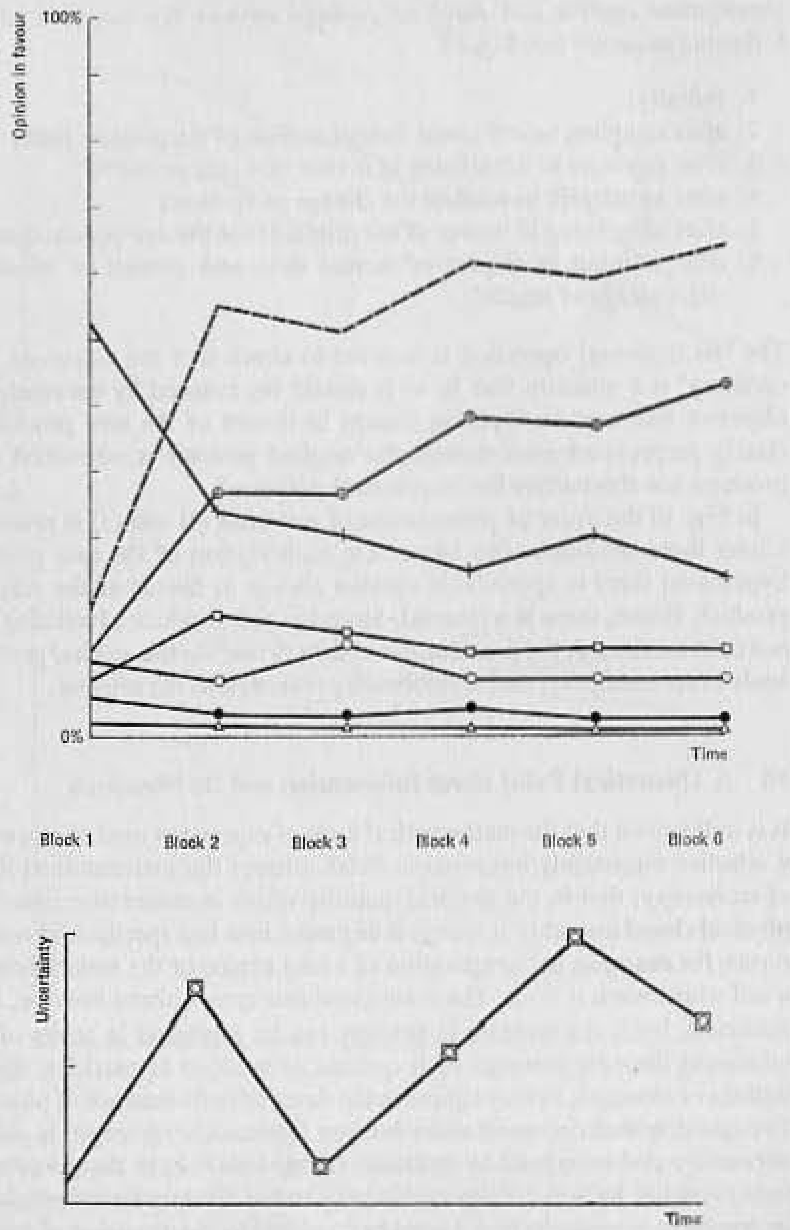



*Figure 9* Forty-seven respondents, six products or installations. Key: ⊕, □, new offers; ○, ×, original offers (same fuel); ●, △, original offers (rival fuel); {⊕ ∨ □}, new product alone or in package (original products are thus (100 per cent − (⊕ ∨ □)).





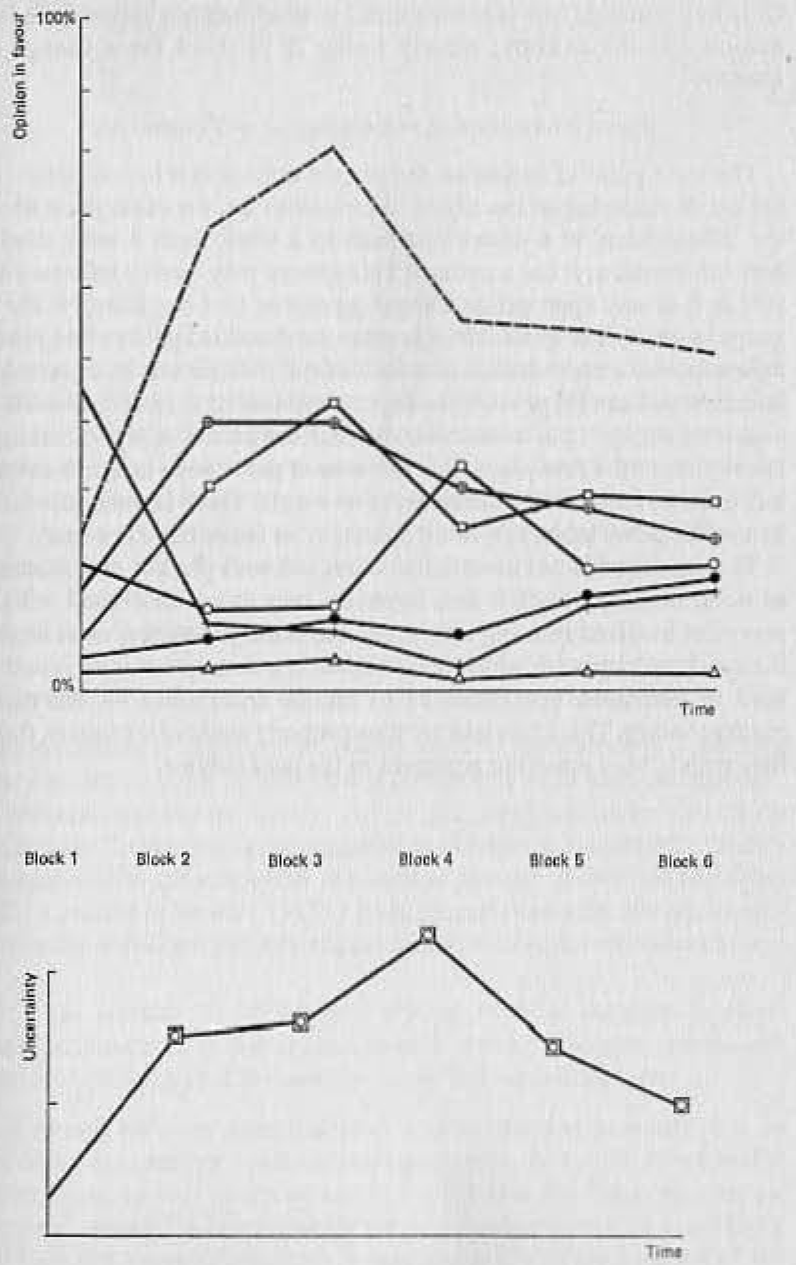



*Figure 10* Seventeen respondents, six products or installations. Key: as shown in *Figure 9*. (Study by: E. Pask and P. J. Mason.)

Only in the former case is there a direct correspondence between the information and the entropy; namely (using $\Delta$ to stand for a change in a quantity)

$$\text{Bound information} = -\Delta\ \text{entropy} = +\Delta\ \text{negentropy}$$

The main point of impact on the present argument is in connection with the act of retrieving or inscribing 'information' as, for example, retrieving the 'information' in a description such as a table. Such a thing does not *have* information; it has a pattern. This pattern may convey information to you or I, or any appropriate computing engine that can interpret the language in which it is written. In this sense the described pattern has *potential information to a user* which is obtainable (a) if the user can interpret what is inscribed and can (b) provide the negentropy needed to read the description and thus engage in a transaction. Similar comments apply to making an inscription in the first place. For this reason pains were taken to calculate information indices over occurrences or events. There is only information in a contingency table, recording events, in so far as there is a user.

The argument is not essentially concerned with physics and mechanics as such. Some parts of it are, however, very much concrened with the processes involved in using information and the coherency, or at any rate the synchronicity, with which information in a desciiption is activated and used to determine operations or to reduce uncertainty on the part of sentient beings. These remarks are thus properly made at the outset, though they mainly bear upon the argument in the next volume.



# 2 Machines

Most engineers think of a machine as a tangible entity such as a lathe or a diesel engine. Naturally they include telephone exchanges and computers as machines also, so that 'machineness' does not depend in an essential way upon energetic transformations; to crunch numbers in data processing is as legitimate a machine activity as crunching rock from a quarry. All the same the engineers' image has a healthy tang of materiality about it and at a later stage we shall retrieve the materiality as a general notion of 'efficiency in embodiment and execution'. But at first it is more profitable to look at the other side of the coin and to emphasise the 'process' inherent in a machine and machine organisation.

## 1 Abstract Machines

In sharp contrast to the engineer, Ashby captures the essence of a machine in the abstract; a point of view that is compatible both with the engineer, and abstract automaton theory. All of the machines under discussion should, until further notice, be regarded as abstract automata but Ashby's treatment of the subject, which the serious student should read (Ashby, 1964b; reprinted in Stewart, 1967) is more general than most and has a built-in guarantee of realisability (find the guarantee, as it is a non-trivial problem).

**1.1** The essence of the simplest type of abstract machine, a state-determined machine, is that it constitutes a 'coding of simple succession'.

The crux of Ashby's definition lies in the following concepts:

1. One and only one state occurs at a time. It being assumed that an observer or designer can point out, independently, the system which is said to have states, so that states are sets of (exclusive and exhaustive) alternatives, $z \in Z$, where $Z$ is the *state set* (of whatever is observed to be a machine).
2. That the constitution of $Z$ is unchanging over the interval of the observation.
3. A state is a complete account of the condition of whatever is observed; if it happens that certain properties were manifest in advance of the states,