Chapter 2
Information and Intelligence

Wisdom is only a comparative quality, it will not bear a single definition

Marques of Halifax,
Miscellaneous thoughts and reflections late 17th century.

2.1 Introduction

The key to understanding intelligence is 'information', since it is information that is the raw material used to gain insight. So we need to appreciate 'information' in a very precise way. The next section will explore a formal definition of 'information' to see if this will help us. It may also give us a different perception of intelligence.

2.2 Information

One of the important consequences of insight is the formation of a hypothesis that has been triggered by a puzzle, as I have illustrated in Chap. 1. In general, hypotheses are propositions that express constraints, laws or rules about the world. From a pragmatic point of view we can say, *hypotheses are useful if they make the world "a less surprising place"* (after Peirce 1958, 1966).

It so happens that independently to Peirce, a measure of surprise had already been derived from communication theory (Shannon and Weaver first printed in the Bell System Technical Journal in 1948 and published in book form in 1964). The problem they were trying to resolve was getting some kind of measurement in communication engineering. This measure should provide a precise assessment and comparison of any non-perfect communication systems. They needed a way of asking, "How good is this communication system?" In particular, they wanted a measure for such systems so that their performance and limitations could be predicted; a measure similar to that of horsepower for engines, where the limits of speed and acceleration can be defined. The model of communication or paradigm they had in mind was the transmission of Morse code. A schematic diagram of the components of a their idea of a typical communication system is shown in Fig. 2.1.

Thus if someone transmits a message that consists of string of digits (or letters) such as:

T. Addis, *Natural and Artificial Reasoning,* Advanced Information and Knowledge Processing, DOI 10.1007/978-3-319-11286-2_2

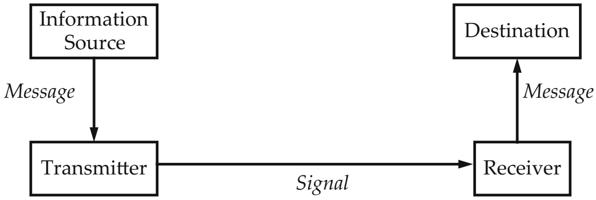

Fig. 2.1 Communication system (Simplified from Shannon Fig. 2.1 1948)

31415926535897932384 6...

where the digits will be converted into a signal by a transmitter, passed down the channel of communication (e.g. a telephone cable), and en route may be changed by the effects of noise (not shown in diagram). The receiver then converts the noisy signal back into digits. From the destination's point of view we should note that:

> The significant aspect (of this communication system) is that the actual final message is *selected from a defined set* of possible messages. (Shannon 1948, 1964).

If we take the simplest notion of a message at the receiver by considering that each digit is a message, then our expectation of a message (i.e. a digit) before it arrives is given by the choice of 1 in 10. So, if each digit were equally possible then the probability that we could guess at the destination point what the next digit would be is 0.1. When a message arrives our uncertainty will reduce to zero because there is now no need to guess (probability is 1). The larger the choice the greater is the initial uncertainty.

We can therefore propose a measure of 'uncertainty' that is inversely proportional to probability: it increases as the probability of choosing correctly decreases. However, this inverse probability measure should also reflect our own perception of uncertainty.

It has been shown that a person's sensitivity to sensations such as hearing or touch is 'logarithmic'. This natural detection system allows us to cope with very loud sounds or firm pressures and yet still remain sensitive to very low sounds or gentle touches (say).

From these observations of choice and the logarithmic scale of sensation, it would be reasonable to define the 'unit' of uncertainty as 0.5, because uncertainty is highest when the probability of either choice is the same. That is when there is zero bias in the choice.

To reflect both the unit of choice and human sensitivity, we can create a function by using the logarithm measure to the base 2 of the inverse probability of a message. The inverse of the probability is used because as the probability of a message rises, the less information it provides. The advantage of this is since a probability is always less than 1 the inverse will always be greater than 1. When a log is taken it will always provide a positive number: (Fig. 2.2)

$$\text{Uncertainity} = \log_2 \frac{1}{\text{probability}}$$

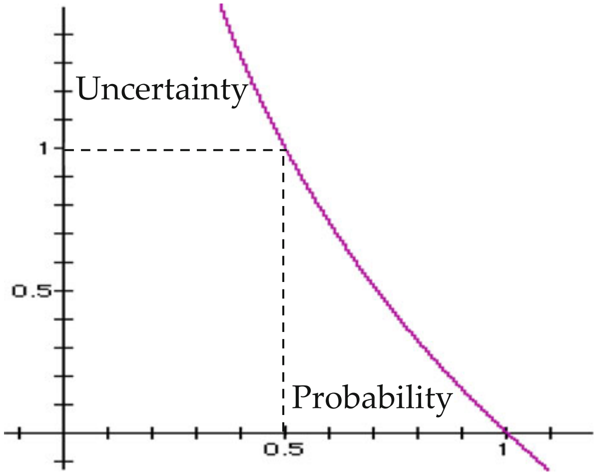

Fig. 2.2 A possible measure of uncertainty where the Unit is probability = 0.5

or

$$\mathbf{Uncertainty} = -\mathbf{log_2}(\mathbf{Probability})$$

On the other hand, if the digits were not equally probable such as in the string:

2222221222222242222232

then we would have a good chance of guessing that the next digit would be 2. In the extreme case, if the digit was always 2 and the system was noise-free, our chance of guessing correctly is certain and no further information is obtainable; that is, the information provided by each message is zero.

- *So we can say by extension that there is more information in a string of symbols where the probability of each symbol is the same than there is in a string of symbols where the probability of the symbols is **not** the same.*

If we imagine the message is coded so that the significant characteristic of the number (the symbol) is whether it is even, odd or prime then our choice is reduced to only three symbols. In this case we would have a better chance of guessing the next symbol (even, odd or prime) than for guessing one of the ten numbers.

- *So there is more information (uncertainty) in a string of symbols where the choice of symbols is higher.*

Finally, the measure of information should have additive properties with a consistent interpretation. So if there was a 1/2 chance that the number is odd and then a 1/3 chance that the odd number is prime (say), the probabilities should combine such that there is 1/6 chance of guessing it to be prime at the start.

- *Thus if the choice to be made is broken down into two or more choices such that the final outcome has the same uncertainty, then the information should be the same.*

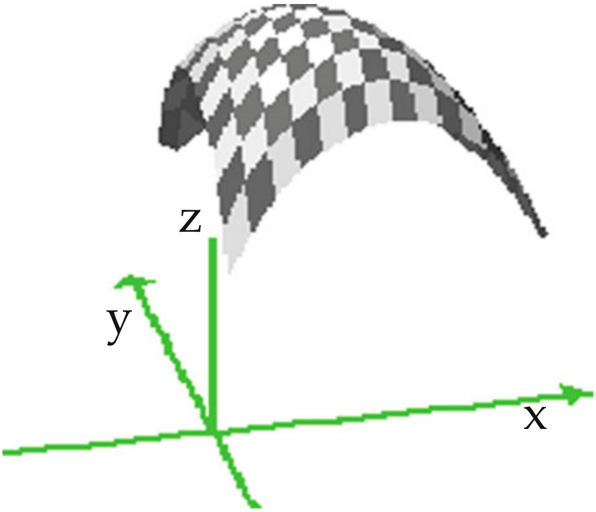

Fig. 2.3 Range of information for three symbols

The only known function that satisfies these requirements is one based upon the expected (similar to average) logarithm of the inverse probability ($\mathbf{p_i}$) of each symbol ($\mathbf{i}$) thus:

$$\textbf{Information of an event} = -\sum \mathbf{p_i log_2 p_i}$$

We can now define the unit of information, which has been called a ‘bit’, where the two choices are equally likely thus:

$$\mathbf{1 = -(0.5log_2(0.5) + 0.5log_2(0.5))}$$

This information measure of a system is called *entropy,* and its behaviour for three choices can be illustrated in Fig. 2.3. In this graph, z represents the information value (in bits) as the probabilities of two (x, y) of the three symbols (w, x, y) are changed. The probability of the third symbol w is determined from the other two probabilities because:

$$\mathrm{w + x + y = 1}$$

The information measure (entropy) falls to zero when the probabilities are 0 or 1, and rises to a maximum for equal probabilities. The maximum in the zx or zy plane is less than the maximum for a plane that includes all three dimensions zxy. The equation for this surface is:

$$z = -\log_2^{-1}[(x\log_2 x) + (y\log_2 y) + ((1-(x+y))\log_2(1-(x+y))]$$

- *So we can say that the greater the entropy, the larger the uncertainty.*

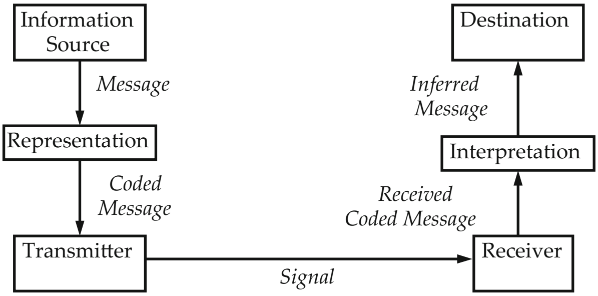

Fig. 2.4 The role of hypothesis in communication

2.3 Insight

So far I have suggested that the message is represented by some characteristic of the transmitted symbols, and this is considered to be a single event. Because of the additive properties of information, the entropy of a sequence of (say) N symbols *that are independent* is the sum of the entropy for each of the symbols.

It could be the case that the sequence of symbols is significant (as in Morse Code) and that each message is identified by a different sequence. If we at the destination 'know' the key of this code, the sequences can be interpreted and the information measure relates to the number of encodable messages (see Fig. 2.4). This will usually be less than the sum of the entropy for the individual and independent symbols.

However, if the key is not 'known', the information is perceived to be that of the uncertainty of the independent symbols rather than the potential messages. This greater entropy we will call *Perceived Entropy*.

- *We can thus say that the perceived entropy is either higher than or equal to the actual entropy of a system.*

In the original sequence of digits above, the probability of any digit occurring is about 0.1, but the insight that this sequence is the value of p means that the sequence of numbers can be calculated from an equation such as:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \frac{1}{13} - \frac{1}{15} + \ldots..$$

- *In this case the perceived entropy falls from that of approaching infinity to zero in a single moment, and it is this insight that characterises the intelligence process.*

Insight can now be seen to involve at least two processes (see Fig. 2.5). The first process is the identification of the symbol (Wittgenstein L 1921). Since the symbol is to be abstracted from the signal we will call this process *abstraction*. The symbol is not always the obvious sign such as a digit, but may be a feature of the sign such as the notions of even, odd and prime. Abstraction may be considered formed through the *perception* of significant features. In this case a perception is a concept that involves bringing together the features into a single unit. That is the identification of the elements observed that carry the important information.

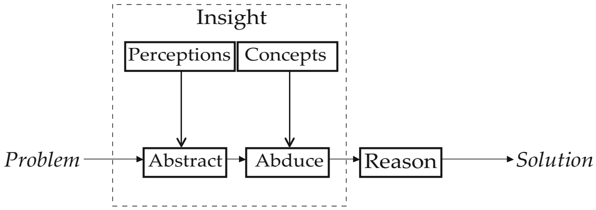

Fig. 2.5 The intelligence process 2

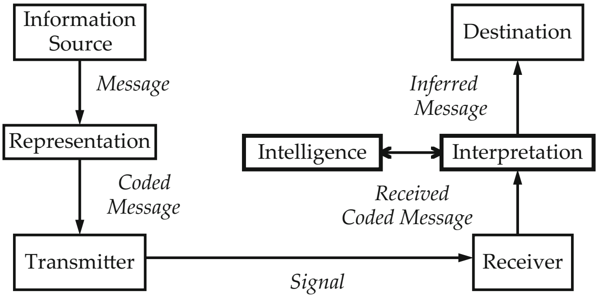

Fig. 2.6 The role of intelligence in communication

The second process is the proposal or generation of a model (an equation in the above case for π). The model must be guessed from the signal, and thus the process of guessing will be called *abduction*. Note that the abductive process does not guarantee a completely successful model. Strictly speaking both these processes are abduction, and to distinguish between them, the second process may be called *retroduction* (after Peirce).

We can now see that intelligence changes the interpreter (the model) so as to minimise the entropy of the communication system (see Fig. 2.6). *We need 'purpose' in order to avoid the trivial solution of minimising of entropy by switching off the signal altogether*. We must assume that the correct identification of the message is important. The problem is that the intelligence process does not know either what the range of messages might be or what part of the signal carries the messages.

2.3.1 The Distinction Between Information and Knowledge

As an aside, it has puzzled some people that noise turns out to have high information because of its unpredictability. We can now see a distinction between *information* and *knowledge* by asking the question of a signal, "*This is information about what?*" Only those events that provide material evidence towards the act of insight and lead towards the reduction of entropy can be called knowledge. Events that are uncertain have varying degrees of information, but *events that are uncertain and contain the seeds of certainty (an insight is possible) represent knowledge*.

2.4 Induction

The process of generating a model and then proving it is useful (i.e. can be used to make predictions) underlies the well known mathematical process of proof by *induction*. For example, if we examine the sum:

$$1 + 3 + 5 + \ldots$$

of the successive odd numbers then we *may* notice that:

$$1 = 1^2$$

$$1 + 3 = 2^2$$

$$1 + 3 + 5 = 3^2$$

$$1 + 3 + 5 + 7 = 4^2$$

and so on. We can *abduce* (infer) the model that *for every natural number n, the sum of the first n odd numbers is* n^2. This is certainly true for all the first n odd numbers from 1 to 4 we have observed so far, and we could continue in this vein until we find an exception. For many scientific endeavours this may be the best we can do but for mathematics this is not considered good enough. Since we have access to the underlying foundations of mathematics the possibility of a sound proof is available.

Such a mathematical proof by induction follows the style: if we provide a general form that shows this model to be true for any number n, then we are entitled to suppose that it is true for any number less than n. We are also allowed to suppose that we already know that the sum of the first $n - 1$ odd numbers is $(n - 1)^2$. The sum of the first n odd numbers is obtained by adding the nth odd number, which is $(2n - 1)$. So:

$$\begin{aligned} \text{Sum of the first } n \text{odd numbers} &= (n-1)^2 + (2n-1) \\ &= (n^2 - 2n + 1) + (2n - 1) \\ &= n^2 - 2n + 1 + 2n - 1 \\ &= n^2 \end{aligned}$$

It should be noticed that:

1. First a proposition (i.e. hypothesis or model) must be proposed (insight). The proposition comes from a set of *concepts* in which each proposition can be used to generate a potential series through deduction. Usually the series here involve sums rather than multiplication. Multiplication series are much more difficult to prove.
2. Then the proposition is tested against observation (reason); a process of *validation*.

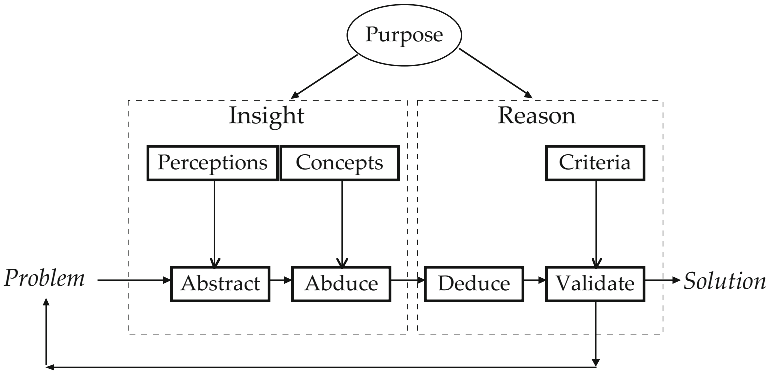

Fig. 2.7 The intelligence process 3

Using a proposition to infer a consequence is the process of *deduction,* and the *deductive* process generates consequences (results). In this case, the sum of odd numbers proposition, we tried out several examples of the series and within these limits the proposition works. However, mathematics does not consider this form of validation sufficient; it has a more exacting *criteria*. This is because, unlike empirical science, mathematics can often produce a general proof that will show a proposition to be true for all possible cases. For such proofs an accepted protocol is laid down. We can now extend our model of intelligence by expanding the process of reason as in Fig. 2.7.

We indicate here, with the feedback loop, that validation can fail and the process will cycle until a solution (of some sort) is found. The decision to finish a cycle depends upon a *validation criterion*. Such a criterion will be different depending on the kind of problem to be solved.

Mathematical induction will thus explore formally every natural number. This can be done through a proof, but if a proof cannot be determined, a search is often performed to find an exception through simple enumeration. Even this latter approach may be too extensive to be practical.

Exactly what *criteria* are invoked to satisfy the validation is left unstated. If a formal proof is not possible (as is the case with empirical science) at what stage do you stop enumerating and testing; when do you just accept the hypothesis? A verse written by a school friend in my wife's autograph book expresses the problem nicely.

With what confusion thinking's fraught,
I often think I'll think no more.
For when I spend much time in thought,
I un-think things I thought before.
Anon

The pursuit of better interpretations of uncertain events characterises intelligence. We have shown that insight is the key that unlocks these interpretations from the events. The question now is, "What are the mechanisms of insight?"

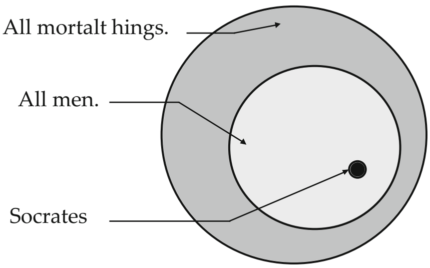

Fig. 2.8 A diagrammatic equivalent of syllogism

2.5 Deduction

Clearly deduction is one of the processes required, since once a proposition is proposed it is needed to create a result from given 'facts'. The normal *deductive* process can be illustrated by the syllogism:

1. **All** (men), **are** (mortal)
2. (Socrates) **is a** (man)
3. **Therefore** (Socrates) **is** (mortal)

The deduction process is a formal procedure that is clearly mechanical, since it does not involve the meaning of the words or symbols (given in brackets) when framed in this structure. The first sentence (the 'proposition') links two phrases together such that the first phrase 1 is said to 'contain' the second phrase 2 as a 'fact'. The first phrase states that a general class of object (men) a share a property (mortal). The second sentence gives an example of the general class (men) as an example of the general class of objects. It therefore follows that this particular example (Socrates) will have this property (mortal); after all, it has just been stated (also see Fig. 2.8).

Deduction contains no uncertainties and therefore does not provide any information. During deduction a marker called the Truth-value tracks the tracing of the certainties. The general form of this deduction is:

1. **All** (A) **are** (B).
2. (X) **is an** (A).
3. **Therefore** (X) **is a** (B).

We can replace the three phrases in brackets by any other statements of facts that we like. If the first two sentences are True after this replacement then the third sentence will also be True since deduction preserves the marker True.

Deduction is a single step in a set of steps that will lead to conclusions that are guaranteed to be True. Consider the following conundrum:

a. Brothers and sisters have I none
b. but this man's father
c. is my father's son
Who is he?

We can choose the following route of syllogisms:

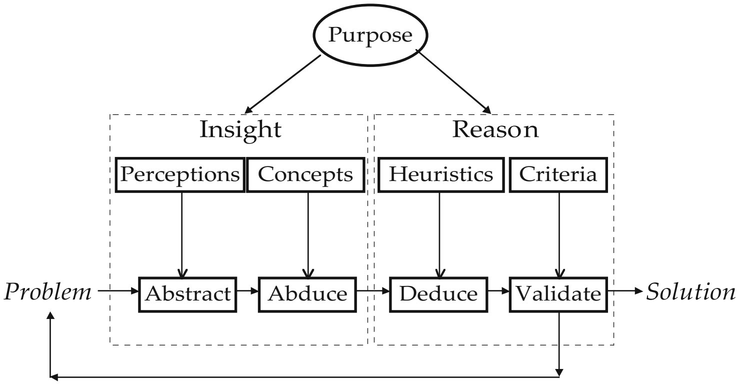

Fig. 2.9 The intelligence process 4

Chose hypothesis(the general case):
All (male who has no brothers and sisters) **are** (the *only* son of a father)
Then given (from line a):
(I) **is a** (male who has no brothers or sisters)
1. Therefore:
(I) **is** (the *only* son of a father)
We infer (from lines b. and c.):
(This man's father) **is** (the son of a father)
And using 1:
(This man's father) **is** (I)
Therefore using a known relationship:
(this man) **is** (my son)

What is not described by this formal layout is why we might choose this particular set of facts to make these particular steps as against the infinity of other possibilities. We never considered the line of daughters, or the many other human relationships that might have been chosen. There is nothing in the rules of deduction that offers guidance to a useful conclusion.

To solve a problem using deduction requires direction; deduction needs problem-solving knowledge that limits the choices amongst the known facts and possible hypotheses. Such problem solving knowledge provides a compass from which to steer our course through a labyrinth of possible steps. This guiding knowledge is known as a *heuristic*, and we should include it in our intelligent process (see Fig. 2.9).

We can now extend the model of the intelligent process to include:

- *perceptions* that identify the combination of features for a useful abstraction,
- *concepts* which are a set of generalisations that can be fitted to abstractions,
- *heuristics* which select the route through to a solution,
- *criteria* which provide the basis on which to accept viable hypothesis.
- *Purpose* that governs all the above.

Springer

http://www.springer.com/978-3-319-11285-5

Natural and Artificial Reasoning
An Exploration of Modelling Human Thinking
Addis, T.
2014, XII, 199 p. 59 illus., Hardcover
ISBN: 978-3-319-11285-5