The Nature of Information

G. G. SCARROTT
34 Parkway, Welwyn Garden City, Hertfordshire, AL8 6HQ

It can be observed that the ultimate function of information in nature is control of the organisation of life. Arguments supporting this proposition are outlined together with some of its consequences.

Received July 1988, revised September 1988

1. INTRODUCTION

During the last few years many of the more perceptive workers in the information systems field have become uneasily aware that, despite the triumphant progress of information technology, there are still no generally agreed answers to the simple questions – What is information? Has information natural properties? What are they? – so that their subject lacks trustworthy conceptual foundations.

The problem was addressed by a set of papers published in the July 1985 *Computer Journal*, but this was only a first step, and there is still a need to put forward a coherent synthesis of current concepts. The general recognition that mastery of information *technology* is necessary but not sufficient doubtless accounts for the recent introduction of the term 'information engineering' into the official glossary, since 'engineering' refers to a useful synthesis of ends and means whereas 'technology' refers only to the means.

The reduced emphasis on pure technology is a step in the right direction, but regrettably the change, for the present, has only cosmetic significance, so that there remains a need to formulate an understanding of the nature of information derived from detached observation of the functions and structure of information in nature. This paper will be an attempt to contribute to such a study without any particular utilitarian objective in mind. If eventually a coherent understanding of the nature of information can be formulated, there is little doubt that it will help to guide the deployment of information technology in useful systems. However a good case can be made for allowing the initial stages of the study to be driven mainly by curiosity regarding the place of 'information' in our understanding of nature at large, and this will be the form of the argument presented here.

2. INFORMATION AND ORGANISATION

A dictionary definition of 'information' as that which informs is true but not very helpful, since it does not address the consequential question, why should anyone want or need to be informed? Sometimes, but not always, people act immediately upon information, so that when this occurs information could be said to control people's actions directly. Quite often people hoard information without putting it to immediate use and, if asked, cannot say for what purpose it will eventually be used. However, if hoarded information is kept for a long time without eventually finding a use by guiding action it is usually discarded. This suggests that the ultimate purpose of information in human affairs is indeed to control physical action, and the other roles of information such as selecting which action to initiate are subsidiary. This raises the question, why is it necessary and acceptable that people's actions be controlled externally? The obvious answer is that a measure of control of people's actions is necessary whenever the people concerned, controllers as well as controlled, belong to an organised social group whose collective purpose is to survive and prosper as a group. This situation occurs very commonly either because their actions constitute the internal operation of an organised commercial company or because they represent the organised interaction of such companies or indeed of society at large. We may conclude that the most basic function of information is to control action in an organised system and thereby operate the organisation.

It remains to formulate a definition of organisation. The essence of organisation is controlled interdependence, but a complication arises from the fact that we use the same word 'organisation' to refer to interdependence at every scale. This can be accommodated by defining organisation and information recursively. *An organised system (OS) is an interdependent assembly of elements and/or organised systems. Information is that which is exchanged between the components of an organised system to effect their interdependence.*

Every living organism, its vital organs and its component cells are organised systems bonded by information within the definition above, and may therefore be termed natural OSs. A social group of living organisms such as a pack of wolves, a nest of ants or a human tribe also has the essential feature of controlled interdependence, so that they also qualify as natural OSs. A deliberately formed group of people such as a commercial firm can be regarded as an artificial OS, as is also every machine conceived and used by mankind, but no inanimate natural object or system qualifies as an OS, since they blindly obey the laws of nature without any control mechanism to avert disaster. Hence the concepts Information, Organisation and Life are distinct but inseparable, since none of them can be defined without reference to the others.

The control mechanism in every living organism consumes energy. Natural selection has tended to minimise the energy level of signals that carry control information, since reduced energy consumption in an environment where energy is the ultimate scarce resource has obvious survival value. The limit to such a minimisation process is set by intrinsic physical uncertainty in the behaviour of very small objects in living control systems. It may be assumed that this limit has been reached in the course of the evolution of life so that the behaviour of every living organism, including mankind, in response to an environmental stimulus cannot be expected to be fully determinate. Indeed this is commonly observed. It is not due to human folly, but it is a consequence of the operation of physical law.[1]

A typical living organism is autonomous, so that it controls its own actions as a consequence of its own understanding of the relevant environment. Thus in a living organism information is understanding, since understanding controls action directly.

Mankind has evolved a powerful technique (natural language) for using symbols to achieve partial communication of understanding between individuals so that they can control one another's actions. This has enabled mankind to operate a flexible, dynamically adaptive social organisation (i.e. a large-scale OS) that has conferred supreme survival value in competition with other species.

Man is the organising specialist of the natural world, totally dependent for survival on his information-handling skills. He therefore uses them for additional purposes other than the immediate control of action. In particular, experience represented by symbols has been accumulated, distilled and re-distilled to constitute knowledge, which is then used to great advantage. For example, knowledge can be used to avoid costly erroneous actions whose effects have already been encountered. Man also uses his information-handling skills as a weapon for social competition by concocting misinformation to foster misunderstanding in his enemies, and in his spare time he keeps his comprehension and communication skills in trim by using information as a universal plaything in the arts and games. Not all misinformation is, however, produced deliberately. Much of it arises from the inadequacies of the human acquisition, storage, retrieval and processing systems and their interaction (or lack of it) with the organisation of which the human is a part.

3. INFORMATION AND COMMUNICATION

On account of the recursive definition of an OS and the information that bonds an OS any statistical patterns in the use of symbols to represent information must apply equally at every scale, so that the resultant overall pattern must be self-similar in the sense defined by Mandelbrot.[2] This argument offers an explanation of the hyperbolic distribution (Zipf's law)[3] for the use of semantic symbols to represent understanding. Similarly Pareto's law [4] for the distribution of incomes can be explained as a direct consequence of the recursively defined structure of organised society.

The popular belief that the refinement of micro-electronic information technology will lead to a new 'Information Society' is an exaggeration. Society has always been bonded by traffic in information, and no other kind of organised society has ever been possible. Certainly the enormous increase in speed and cost/effectiveness offered by micro-electronic technology opens up new possibilities for larger-scale organisation and faster response to external stimuli. However, not all such possibilities are necessarily desirable, since human response time is unaffected by technological advance so that faster communication may simply make a large-scale mixed organised system of people and machines more unstable, and indeed this has recently occurred. Nevertheless, there is always a requirement to improve the quality of human judgement by improving ease of access to potentially relevant information. There is no doubt that refinement of micro-electronic technology will facilitate such developments, provided that it is recognised that the design objective is to aid human judgement and not to provide a substitute for it.

Information, as defined above, can be observed to operate in every living organism. To appreciate the nature of information in human affairs it is essential to distinguish clearly between understanding, which directly controls human actions, and the symbol combinations used to make a partial representation of understanding for communication. Only the latter are accessible for use in information machines. They are conventionally known as 'Data', but are too often confused with 'Information', i.e. Understanding. It is now, when we are attempting to introduce 'understanding' into Data Processing, that we must understand clearly the nature of the medium with which we are dealing.

The classic work by Shannon and Weaver entitled *The Mathematical Theory of Communication*[5] was concerned only with the reliable communication of symbols and made no claim to be a theory of information. Unfortunately it is often referred to in the academic world as 'Information Theory', and this has inhibited general recognition of the need for a theory of information.

The concept of entropy as defined by Shannon is essentially entropy communicable by symbols, since he was formulating a theory of communication, and he made no claim that all the symbol combinations that contribute to the entropy were necessarily meaningful in users' terms. Shannon's communicable entropy is ultimately based on a Markov process for generating symbol strings, in which the choice of each symbol is statistically correlated only with its near neighbours. This simplifies the mathematics and is quite satisfactory for tackling the engineering problem of communicating symbols in the presence of noise, but a finite Markov process cannot model symbol strings generated by the recursive process used to construct meaningful text, in which correlations occur at every scale. Consequently the established entropy measure in bits, for example of storage capacity, is not a valid measure of the utility of the store in human terms. Hence widely used measures of technological cost effectiveness such as 'cost per bit' are invalid. Indeed this proposition is obvious, since for example the phrase 'cost per bit' applied to a book would imply that the value of a book in users' terms is proportional to the number of letters in it and consequently to the weight of paper, an absurd conclusion. Certainly the production cost of a book may be proportional to the number of letters in it, but its intrinsic value in users' terms is a more subtle matter. To tackle this problem it is necessary to define usable entropy (UE) of a collection of symbols with communicable entropy (CE), as the binary logarithm of the number of meaningful texts that can be constructed from them. It is obvious that the UE of a collection of symbols is very much less than their CE, since for example if the words of a sample of meaningful text are jumbled at random most of the resultant word combinations are not merely wrong, they have no meaning at all. Hence the hyperspace with dimensions defined by CE is very sparsely populated with meaningful points, a natural property of information that is put to good use in hashing techniques.

Evidently the UE of a collection of symbols is a function of their CE. The function cannot be expected to be precisely defined on account of the intrinsic disorder in the human use of information, but we need to define it approximately to permit a realistic commercial comparison of competing storage devices to be made in terms of their capacity Q, access time T, and cost C. When a typical document is created, to ensure that it is meaningful, each addition must not contradict what has already been written, so that the author's freedom to select new symbols is progressively reduced as he creates the document. Similarly, a large collection of documents must be subject to internal correlations to avoid gross contradictions, so that the symbols that represent them cannot be chosen independently, as assumed in the calculation of communicable entropy. Hence as we increase CE, UE also increases but much more slowly, so that the ratio UE/CE is reduced, ultimately to zero as CE tends to infinity. Hence the technological cost effectiveness of a storage device is $F(Q)/CT$, not Q/CT as commonly assumed, where function F has the property that as Q tends to infinity, $F(Q)$ also tends to infinity but $F(Q)/Q$ tends to zero.[6]

A plausible derivation of the nature of F can be deduced from the conjecture that the usefulness in human terms of a store is proportional to the number of recursion levels in an organised collection of symbols that it can accommodate. It can be observed that the recursively defined tree structures that are so common in human affairs (for example the structure of Roget's *Thesaurus*) have a well-defined average fan-out at each level of recursion, so that the number of terminal items in a typical structure is an exponential function of the number of layers of recursion in it. Conversely the number of levels is proportional to the logarithm of the number of symbols, so that UE is proportional to log CE. Certainly the conjecture leading to this result is speculative, but the conclusion accords with the common experience that dazzling achievements in the development of storage technology have often produced only marginal improvements in the satisfaction of the system user.

4. INFORMATION AND JUDGEMENT

A typical information link in an organised system operates in three phases – observation, judgement and action – since it is only in rare simple situations that observation can be used to control action directly. In living organisms the brain has been evolved to carry out the judgement process by taking into account, against a background of experience, observations from several organs simultaneously, for example sight, scent and hearing, in order to make an appreciation of the situation appropriate to guide action. This process may be termed 'primitive judgement', since it is common to many species. Mankind uses primitive judgement on a grand scale in daily life, but as a by-product of the evolution of natural language the technique of logical argument by the use of determinate operations on the abstract concept of certainty has been added comparatively recently to our human judgement skills.

Computers are essentially logical argument machines, and they are of great value for assisting or undertaking any judgement that can be resolved into logical argument, but since primitive judgement preceded logical argument in the evolution of human information-handling skills, the widely held belief that every judgement must be describable in logical terms is ill-founded.

A subsystem for effecting judgement may be termed an 'Information System' (IS) to distinguish it from an OS, since no action is involved in an IS. Then two distinct kinds of IS can be observed to exist, natural ISs in living organisms and artifical ISs such as computers. A natural IS is constructed from large numbers of slow, intrinsically untrustworthy elements (see Section 2 above) and indeed it is slow and untrustworthy when used to formulate and follow a complex logical argument. However, a natural IS enables a living organism to achieve formidable feats of real-time physical control, such as a man playing tennis or a humming bird hovering, and also makes possible the recognition of complex patterns in the presence of almost overwhelming noise.

An artificial IS is constructed from fast logical elements, whose operations can be assumed to be totally trustworthy, either because they are intrinsically determinate or because tactical devices such as error-correcting codes are used in the system to give the effect that the elements are totally trustworthy. A typical artificial IS can simulate a logical argument by manipulating symbols with much greater reliability and at a speed enormously higher than a natural IS, but it cannot compete with a natural system on pattern-recognition tasks and offers only clumsy techniques for real-time control of physical action.

5. INFORMATION AS AN AREA OF STUDY

Many of the potential applications of logical argument machines (i.e. computers) have already been achieved, but there remains an unsatisfied requirement for artificial information systems to aid human judgement in circumstances where the relevant logical argument is not known or may not exist. Living organisms demonstrate that it is possible to achieve useful judgement of complex situations by methods other than logical argument, for example understanding human speech in a noisy environment by making effective use of available redundancy. Information systems in living organisms are not programmed, but they learn from experience. Certainly they are error-prone, but over a wide range of pattern-recognition tasks, diligent attempts over a long period to simulate human performance by the use of a conventional computer have failed. A clear case can be made for deploying research effort to identify the techniques used in a natural IS and endeavouring to learn from them in the design of an artificial IS, without the constraint that the artificial IS should resemble a conventional computer. Indeed, work already in hand on neural nets is of this nature.

Computers were originally conceived by mathematicians. Consequently an artificial information system has always been regarded as one or more automata that obey a program of instructions that are defined in essentially logical terms, and it has been tacitly assumed that there is no limit to the application field of such a machine provided that the program is conceived and written in a properly disciplined way. These assumptions need to be reconsidered. Information is a by-product of life, and the behaviour of living organisms is not fully determinate for basic physical reasons. Mathematics is essentially formal reasoning, with primitive judgement consciously excluded. Reasoning is the most recent of our human information-handling skills, but it is not their foundation, so that we cannot assume that every judgement of practical importance can be mechanised as a logical argument.

The term 'Informatics' has recently been introduced to refer to all aspects of information, the role of information in human affairs as well as the various techniques used to build useful information systems. Then the relationship between informatics and the more familiar disciplines can be summarised as follows.

Mathematics is a part of informatics that is a part of anthropology that is a part of biology that is a part of physics.

Perhaps this rather clumsy sentence also illustrates the recursive framework of understanding. The proposition that mathematics is a subset of physics may shock some mathematical physicists. Certainly the mechanics of mathematical reasoning are essential for formulating physical arguments, but physics – concerned with the directly observable world – is more fundamental than the abstract world of mathematics, a comparatively recent intellectual construct.

Many of the statements in the above argument are individually obvious and apparently trivial. Their significance arises from bringing them together in a coherent argument concerned with the nature of information. This is seldom done, perhaps as a consequence of the position of mankind in nature as a specialist in information/organisation. We owe our existence and survival to our information-handling skills, so that we use information compulsively and naturally on a grand scale, and we find it difficult to achieve the sense of detachment required to put our own information practices and skills into their proper evolutionary context. Indeed, for a man to ask himself why he uses information is perhaps analogous to asking a bird why it flies.

6. CONCLUSIONS

Many of the points made in the outline argument above can also be regarded as conclusions. The following key points are selected for brevity so that the remainder of the argument follows from them.

1. 'Information' in its most fundamental sense can be defined only in the context of an 'Organised System' (OS). *An organised system is an assembly of interdependent components whose physical activities are controlled so that they complement one another and thereby create a coherent whole. Information is that which is exchanged between the components to effect control within an organised system.*

In a typical OS many of the components are themselves OSs, so that formally it is necessary to define an OS and information recursively. *An OS is an interdependent assembly of OSs and/or elements. Consequently information also has a recursively defined structure.*

2. The essence of life is organisation, since inanimate physical systems do not exhibit the active interdependence that characterises organisation. Hence information and organisation could not exist without life.

3. The inanimate physical world is characterised by the interplay of order and disorder, where disorder implies intrinsic unpredictability. Living organisms are also characterised by the interplay of order and disorder, where the order is recursively defined and where the disorder arises from physical disorder amplified by the control mechanisms of life. Consequently there is an element of intrinsic disorder in human behaviour that sets a limit to the application field of information machines.

4. Evidently living information systems operate very satisfactorily in spite of the fact that they are built from intrinsically untrustworthy elements. We do not know how they work, but a strong case can be made for deploying research effort to find out. The research workers should be motivated by scientific curiosity, without a specified utilitarian objective. Nevertheless, if they are successful there is little doubt that the resultant understanding will be of great value for tackling problems such as pattern recognition and perhaps putting to use some future 'pico-electronic' technology that may be intrinsically untrustworthy for basic physical reasons on account of extreme miniaturisation. A coherent understanding of the nature of information could also serve to discourage the adoption of intrinsically unachievable objectives on the frontiers of information engineering.

5. The finite Markov process underlying Shannon's classic analysis is not a valid model for producing meaningful symbol strings. There is a need to reconsider the calculation of usable entropy for symbol strings constructed by the recursive process, in order to provide a measure of the intrinsic value in users' terms of components such as stores.

6. The computer pioneers recognised the theoretical existence of uncomputable numbers but nevertheless referred to their creations as 'Universal', and this view is still cherished by some computer theoreticians today, but the implication that the potential application field of computers is unlimited needs to be revised. The concept of absolute certainty is a human conceit, valuable, indispensable, but none the less artificial and contrary to natural law, referring only to abstract concepts such as number. It is the foundation of all our concepts of logical argument and the computing machines we have built to follow logical argument. We all share an honourable ambition to justify our judgements by logical argument, but since primitive judgement preceded logical argument in the evolution of human information-handling skills we cannot assume that every judgement of practical significance can be rationalised as formal reasoning. Hence the potential application field of computers, although large as the pioneers recognised, is not universal, and there is a need to explore other forms of information system that are not based on explicit logical argument.

Acknowledgement

The formulation of these arguments and conjectures owes much to the encouragement and constructive comments of Professor J. F. Coales, who has independently recognised the need for a coherent understanding of the nature of information.

REFERENCES

1. G. G. Scarrott, Physical uncertainty, life and practical affairs. *Journal of Information Technology*, **3** (1), 58–61 (1988).
2. B. Mandelbrot, *Fractals, Form, chance and dimension.* Freeman, San Francisco (1977).
3. G. K. Zipf, *Human Behaviour and the Principle of Least Effort.* Addison Wesley, New York (1949).
4. V. Pareto, *Cours d'economie politique*, vol. 2, section 3. Lausanne (1897).
5. C. E. Shannon, A mathematical theory of communication. *Bell Syst Tech J 27*, 379–423, 623–656 (1948).
6. G. G. Scarrott, Some consequences of recursion in human affairs. *Proc. I.E.E.* (part A), **129**, (1) (1982).

Announcements

8–10 November 1989

First International Conference on Computers and Safety, Cardiff

The First International Conference on Computers and Safety will be held at St David's Hall, Cardiff, UK from 8 to 10 November 1989. The Conference, which is being organised by the Institution of Electrical Engineers (IEE), will focus on recent developments and future trends in the design, assessment, application and use of all types of programmable electronics and other complex devices used in safety-related systems.

Increased use is being made of computer-based systems for both control and safety functions. This trend is accelerating at a dramatic rate, and with the acknowledged shortage of experts, many companies are having to rely on suppliers to install such high-technology equipment.

It is realised internationally that the systems are becoming more and more complex and that we are installing technology which is becoming difficult to understand and to assess. It is therefore critical that engineers who are involved in the design, installation, application, assessment and maintenance, etc. of such systems should have a thorough understanding of the underlying principles which affect them.

Topics may include:

Approaches to safety in:

- manufacturing (e.g. robotics/CNC machine tools/drive systems);
- process industries (e.g. emergency shutdown systems);
- electromedical equipment;
- power generation (nuclear and conventional);
- nuclear processing.

Software engineering

- Determination of integrity levels
- Techniques for achieving high-integrity software
- Practical methodologies for industry

General topics, including

- Artificial intelligence in industrial control
- Industrial programmable controllers in system design and assessment
- Guidelines for system design and assessment
- National and international standardisation
- Current and future legislation
- System design techniques (fault tolerant/error detection/recovery/error avoidance)
- Safety-related system architectures
- Qualitative and quantitative approaches to system design and assessment
- EMC considerations in design, installation and immunity testing
- Assessment of communication systems
- Human factors in safety-related systems
- Alarm system management

For further information contact:

Christina Dagnall, IEE, Savoy Place, London WC2. Tel: 01-240-1871 ext. 272.

February 1990

First Call for Papers
Studies in Technological Innovation and Human Resources (Vol. III)
Management and Technology-mediated Communication

'Studies in Technological innovation and Human Resources' is a biannual series of books, published by de Gruyter (Berlin and New York), which brings together research, critical analysis and proposals for change in this fairly new and highly important field of inquiry: technological innovations and how they affect people in the workplace. Technology includes computers, information systems, telecommunications, computer-aided design and manufacturing, artificial intelligence and other related forms.

The first volume of *Technological Innovation and Human Resources* dealt with *Managing Technological Development: Strategic and Human Resources Issues* and was published in 1988. The second volume, entitled *End User Training*, is scheduled to appear at the beginning of 1990.

The upcoming Volume III, *Management and Technology-mediated Communication*, will particularly include papers that are:

- international
- interdisciplinary
- theoretical
- empirical
- macro
- micro

Authors are encouraged to think creatively about subject-matter possibilities, including manuscripts that:

- are critical and reflective;
- report on cases of successful technology-mediated communication in organizational settings;
- report on failures;
- attempt to demonstrate change in communication content, patterns and quality due to technology;
- investigate how speed and complexity may be affected by technology-mediated communication;
- compare various communication means such as telephone, computer mail, video and typewritten and other communication processes and outcomes;
- assess efficiency, costs, benefits and effectiveness;
- combine some of these issues.

Each manuscript must have a concluding section entitled 'Implications for Research and Management.'

Publication procedure

Papers should conform strictly to the rules of the APA (American Psychological Association) style guide (3rd edition). All submissions must be original work which has not appeared elsewhere, and which is not being considered for publication by any other outlet at this time. Since the review process will be anonymous, please prepare your manuscript accordingly. Five copies should be submitted to the address below. Deadline for submission is 1 February, 1990.

If you would like to discuss your topic, please call Urs E. Gattiker at (403) 320–6966 (mountain standard time) or Laurie Larwood at (518) 442–4910 (Eastern Standard Time).

Please submit **five copies** of your manuscript to:
Urs E. Gattiker, Technology Assessment Research Unit, School of Management, The University of Lethbridge, Lethbridge, Alberta, Canada T1K 3M4. Fax: (403) 329-2022; E mail:
GATTIKE%VP@UNCAEDU.BITNET.

5–9 November 1990

ICCC–90 New Delhi, India

The next conference of the International Council for Computer Communication is being held in India in 1990.

For further information contact:

Ms Saroj Chowla, Organising Secretary, ICCC–90 Secretariat, CMC Ltd, 1 Ring Road, Kilokri, New Delhi 110 014, India. Tel: 6830087, 6840052. Telex: 31-66082 CMCG IN.