REVIEW AND SYNTHESIS

B. C. Brookes

1. What is science?

During our discussions it became apparent that we used the term *science* in different and confusing ways. What science?

First consider *pure science*. Science is the body of knowledge about the world we live in which has been found by the activity we call *research*. Note that in considering the nature of science or research we are in no way concerned with the *formalisms* or *techniques* we have discussed. These are merely means to an end - a set of man-made tools from which we each make our own selections when we do research.

When Francis Bacon formulated the Method of Induction 350 years ago he started a philosophical debate about research which has not yet been concluded. The method of induction requires the enquirer patiently to cumulate as much data as possible about the phenomenon under study until the enquirer finds it possible to make some *generalization* or formulate some explanatory *theory* about the phenomenon. The Method of Induction is much used in everyday life in much the same way as Molière's Monsieur Jourdain found that he had used prose. But Bacon was writing in the early 1600's before modern science had emerged.

The philosophers' search for the secrets of scientific method was made more difficult by the astonishing success of Newton's theory of gravitation (1687). It seemed that the theory of gravitation was *true*. Even the great philosopher Kant (1787), who was sceptically critical about many other matters, believed that the Newtonian theory was true. So Kant led later philosophers to consider how the Method of Induction could lead to Absolute Truth. This problem was studied for more than 100 years - until Einstein, with a new theory, explained some of the baffling anomalies that

Newton's theory could not explain. In so doing, Einstein proved that Newton's theory was *not* absolutely true.

Karl Popper (1959) made a decisive break-through. He pointed out that no scientific theory can claim to be True (*vide* the history of science) but is merely the best we have at any one time. All scientific theories, however strongly established, remain always open to *falsification* in the light of later evidence and thought. Popper also pointed out that, in any case, Newton had not derived his theory by the Method of Induction and that there is an asymmetry in the logic of the verification and falsiification. To *prove* that "all swans are white" one would have to observe all "swans": to *disprove* that "all swans are white", it would be enough to produce merely *one* non-white "swan".

Though the logic of Popper's argument is sound, the falisification of a well-established theory is not, in general, achieved quite as simply as he claimed. It is as important to defend older theories against new heresies as to consider what advantages the new heresies may offer.

The debate has been continued more recently by Kuhn, Lakatos and Feyerabend. In opposition to all other philosophers of science, who have been seeking for 'law and order' in science, Feyerabend argues that there is no method at all - that "anything goes". But he also admits that he is being provocative.

My own view is that the scientific method involves *some* 'law and order', ie *some* discipline. In the natural sciences a phenomenon is studied, observations are made, data are collected only until an idea, a 'hunch', a generalization, a hypothesis arises. How? It does not matter. Nor can rules be given. If the hypothesis looks good, it should be tested *destructively* before publication (though referees help in this work). If the hypothesis is published, publication in a scientific journal carries the implicit challenge to all it may concern: refute this if you can! If it is

not refuted, it is criticized, modified and punched into better shape until, in turn, it is knocked out by some better theory.

There are disciplines to be learned. Scientific papers have a distinctive rhetoric. To be taken seriously, the contributor has to use the current terminology of the subject and demonstrate that he knows its recent theoretical history. He has to recognize what constitutes relevant observational evidence and the various forms of fallacious reasoning. He has constantly to refer to the phenomenon under study and to compare *his* description of what he sees with the descriptions of the same phenomenon given by others.

2. Pure *v* applied science and research

There has been no adequate analysis of *applied* research partly because much of it is not published in *document* form: it often emerges as *hardware*.

The main difference between pure and applied research is that, in *applied* science, the search is not motivated by the desire for understanding and truth but by the need to make something "work". A comparison:

PURE RESEARCH	APPLIED RESEARCH
Search for truth and understanding	Search for something which works
Must be published and invite public criticism	Need not be published (except as a patent).
Usually long-term	Usually short-term *ad hoc*
Seeks *generality*	Seeks *particular* objectives
Must take risks: hypothetical	Must "play safe": positivistic

Otherwise, both forms of research may call on the same formalisms and use similar working techniques.

3. Synthesis

(a) The research levels

If the above distinctions can be accepted, at least provisionally, then how do they apply to information science? The unchallenged research areas of information science centre on the operation of IR systems, the needs of users, the organisation and dissemination of information and the various technical processes involved in such work. In all these areas research is already separable into pure and applied. The Forum did not call for papers on particular systems, ie on *applied* research, but several papers on *pure* research were presented, of which Heine's model of IR systems was just one example.

So I conclude that information science has now discernibly separated out into at least two levels of research in which problems which arise directly from the traditional operational areas of the subject are studied.

There was also some evidence that a more fundamental level of empirical research may be arising around the central concept of the nature of information. These two levels of pure research will, in due course, feed back ideas and results into the operational level.

Thus information science now has the research pattern of any other science; it has arrived as a *science*.

(b) The need for coherence

The pure research levels are still very thinly manned and the few active individuals are widely scattered. It does seem, however, that the pure research topics are almost ready to be linked into a coherent corpus of theory so that an integrated *theory* of information science, identifiable and autonomous, will soon emerge. The fact that the research ideas are not yet coherent does not make our subject less of a science: even theoretical physics is not coherent yet.

The operating level is very busily occupied at present with the extension and improvement of information systems and has little interest in pure research. On the other hand, pure research has so far fed back little that could be applied. There is a need for more interaction between the research levels.

(c) Related subjects

Other papers presented at the Forum, for example, that by Samuelson, showed information science interacting with related subjects. This kind of interaction, too, with general systems theory, decision theory, cybernetics, computer science, ...is also important for the development of information science now that it is strong enough to retain its identity. At present, information science is still a net importer of ideas from these other disciplines and it will not be well regarded elsewhere until it has ideas and theories to export to others.

(d) Integrable topics

One of the main possible integrations could embrace the mathematical/statistical models of IR systems with the analyses of dissemination and other explorations of the use of scientific literature. One of the needs is to link these processes with the time dimension and to generalize them into a unified general theory of human communication, largely statistical, which Goffman has already initiated. The time dimension is needed because though information can now be transmitted almost instantaneously, social acceptance of new ideas takes time even in the pure sciences.

A second area which calls for closer interaction centres around various linguistic matters with which information science is involved. Theoretical linguistics has so far been wholly unhelpful in application to information work but in a more closely integrated information science those doing research on linguistic problems would have access to linguistic data of a new kind. The problems of indexing, abstracting, and compaction of titles, etc. offer a new kind of

relationship between language and its use and therefore an opportunity to study language in a new empirical way.

(e) Shannon information theory

The paper presented by Lynch reminded us that Shannon information theory is relevant to some of our problems. In the design of information systems we are concerned with many different kinds of input, many different kinds of user and many different types of search question. All this implies that the system parameters can only be statistical and that Shannon theory should therefore have further applications.

For example, in the *dissemination* of information there is a conflict between what appears to be a natural *scatter* of both sources and users and the need to organize and compact information relevant to any given theme. It is interesting to note that the entropy (in the Shannon sense) of a system of given mean "power" is a maximum when the distribution of sources is gaussian. Gaussian distributions occur in the Swets model of IR systems. The cumulative form of the Bradford law also has the form of a Shannon entropy. At this statistical level of the analysis of information systems and processes the Shannon theory may yet become one of the integrating formalisms we shall need.

(f) The concept of information

There is also a developing concept of *information* itself, the study of which might become the focus of a more fundamental level of research. In the *scientific* exploration of information one would expect empirical detailed studies of individual 'users' before and after they have received some information. In the analysis of this kind of process a different approach is needed in which *structure* rather than *quantity* has to be considered. Such work offers the prospect of a new theory of semantics, based on structural relationships of a new kind, being established from within information science. Belkin has started work on these ideas.

(g) Social aspects

At the present time information science is still mostly concerned with the developments of information systems and networks for scientists. But the field of application widens as new data-bases and networks are established. It therefore becomes increasingly necessary to consider the social implications of the impending 'information revolution'!

Leimkuhler discussed the implications of a wider dissemination of information for the *firm* which (from a theoretical point of view) offers a clearly-delineated entity interacting with its environment - an "open system" - and so provides a test-bed for observational study. Here the main objective of the firm is to ensure its *survival* in an increasingly competitive environment. Kochen's WISE project is concerned with the wider dissemination of information to society at large. And Brookes outlined an evolutionary theory in which man could be regarded as the latest stage in a series of animal species evolving towards an ever increasing information-processing capacity in a struggle for survival on Earth. The unique step that man has taken, through records, reprography (wide sense) and the computer, has been to provide himself with an exosomatic information storage and processing capacity of unlimited scope - a kind of *social* brain which, in principle, can be shared by all mankind. He saw this as likely to be increasingly concerned with the problem of ensuring the survival of the human species - either on Earth or on some other planet.

The continued development of information services is going to raise issues of increasing social importance which need series and early consideration. What is it all for?

(h) The place of the discipline

My answer to the problem considered by Wersig: What *is* information science, and which is implied in my comments above, does not conform exactly with any of the solutions - "broad", "medium" or "narrow" - which he propounds in his

excellent analysis. It seems to me that Wersig's three solutions can be regarded not as mutually exclusive but as representing phases in an evolution from the narrow front from which information science emerged to a broader front as the subject develops. In fact, the whole corpus of active information scientists is always likely to constitute a group accommodating a wide range of individual interests and attitudes, so that all three of Wersig's "solutions" could co-exist.

In this early developmental phase of the subject a diversity of approach could be creative, as Wersig suggests. The issues that separate us are not crucial and will be decided by history rather than by debate.

4. Main Conclusions

(a) The shape of an integrated information *science* with the operational and research levels characteristic of any other empirical science is now discernible.

(b) The central topics of the sciences are information organisation, dissemination and retrieval - the topics unchallenged by any other discipline.

(c) The pure research levels are still thinly manned and the individuals concerned are too widely scattered for the critical interaction which is needed for 'take-off'.

5. Recommendations

(a) We need no longer examine ourselves introspectively. The next Forum on research could more usefully concern itself wholly with specific research topics and their integration.

(b) Those with academic responsibilities need to work towards a research-oriented integration of the subject. We need recruits to the pure research levels.

(c) Theoretical research in information science is still marked by a tendency to 'play safe' and to be positivistic because it has grown only slowly out of *applied* research. So it is still marked by timidity. It could now afford to be more boldly speculative, intellectually exciting and therefore more attractive to intelligent and ambitious students.

---oOo---