Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

Brookes equation: the basis for a qualitative characterisation of information behaviours

David Bawden

Department of Information Science, City University London, UK

Abstract

Brookes equation, which has hitherto been regarded as a kind of non-mathematical shorthand description of information use, is extended so as to show the qualitative pattern of information access. Information inputs are represented by a simple power law, and changes in knowledge structure by possible changes to a simple mental map. The resulting pattern shows seven forms of information input and three forms of change in knowledge structure, giving twenty-one categories overall. These have potential for use as an explanatory tool and taxonomy for studying and summarising information behaviour, and offer the possibility of being developed into more sophisticated and quantitative treatments.

Keywords: Brookes equation; cognitive paradigm; power laws; mental maps; information behaviour

1. Introduction

In 1980, Bertram Brookes finalised a line of research which he had followed for several years, by publishing, in a paper in *Journal of Information Science*, what he described as 'the fundamental equation of information science' [1]. This equation, which he had put forward in various forms, was here presented as:

$$K(S) + \Delta I = K(S + \Delta S)$$

A knowledge structure, K (S) is changed into an altered knowledge structure K (S + ΔS) by an input of information, ΔI. ΔS is an indicator of the effect of the modification. The same information may have different effects on different knowledge structures. It is a description of the information communication process as it affects one individual's knowledge, and hence is regarded as a foundation of the cognitive paradigm in information

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

science [2, 3, 4, 5]. Cole used it as a "baseline" for a consideration of a subjectivist construction of information [6], while Ingwersen (1992) derived extended versions of the equation [7].

Brookes himself, and numerous authors subsequently, have pointed out that this is not an equation in the usual mathematical sense; Todd [8] and Cornelius [9], in reviewing the value and applicability of the equation, give accounts of these viewpoints. The equation as presented gives us no way of quantifying the amount of information of knowledge in K (S), nor any way of deriving the difference – qualitative or quantitative - between it and K (S + ΔS), although Cole has shown one way in which this might be done [10]. Brookes acknowledged that it was pseudo-mathematical and over-simple, and that its terms and symbols were undefined. Neill [11, p. 36] suggested that just the undefined "+" sign "contained an entire discipline"

Despite this lack of equation-ness about Brookes' equation, it has been often quoted as a qualitative summary of a central issue within the information sciences; the Thomson ISI databases in mid-2010 showed the paper having been cited over 100 times. As Cornelius says "It may have been unoperational in information systems, but it has remained operational as a general consideration, even if not in experimental design, within information retrieval theory and within information science's theorizing of information" [9, p 407]. It has also been applied to a very limited degree in practice: qualitatively, as a framework for the study of general public information utilization for health information [8, 12], and quantitatively in a study of archaeological research [10]. However, "for the most part… the equation remains as a rather reproachful reminder of how far we are from fulfilling Brookes' research programme" [13, p 419].

It seems therefore that it may be worth examining the equation further, to see if, by making some, hopefully reasonable, assumptions, it may be possible to go beyond the very limited, almost pictorial, use of the equation which has been the norm. Specifically, the purpose will be to see if the equation, thus modified, will yield a sensible and realistic qualitative picture of information transfer. If so, this would suggest that further studies might offer the prospect of quantitative study.

First, in part 2 of the paper, we shall examine the ΔI term in the equation, the information, or knowledge, input. Then, in part 3, we shall examine ΔS, the change in knowledge structure. We will consider whether the equation can help to give a qualitative understanding of *what kind* of information input is occurring in any particular case, and *what kind* of change in the knowledge structure in occurring. In part 4, we shall combine these, to give a qualitative categorisation of the change described by Brookes' equation.

2. ΔI: a small bit of knowledge

Brookes described this term in the equation as represented an input of "a small bit of knowledge" [1, p 131]; however, he preferred to term it information, to make the point that this had an objective quality, albeit that it might cause different effects on the knowledge structures of different individuals.

Brookes presumable used the word 'bit' in its informal meaning; but we could, of course, quantify any information input in terms of bits, in the information theoretic sense. This would not, however, help in understanding what kind of information input it is.

To begin to do so, we will treat ΔI not as a number, but as a function. We do not know, *a priori*, what this function is; not even its general nature. But we may take an educated guess that the most likely form of such a function would be that of a *power law*. This seems likely, simply because power laws are very commonly found in many aspects of the biological and social domains; it is difficult to see any rationale for choosing any other form of function. Within the information sciences, they are encountered as rank-frequency functions, such as Zipf's Law and Pareto's Law, and as size-frequency functions, such as Lotka's Law [14].

These take the general form

$$y = a\, x^{s}$$

where a is a constant and s is the exponent of the function.

This results in a series where the value for the k^{th} element is

$$\frac{1/k^s}{\sum 1/n^k}$$

the denominator being summed over all the terms of the series.

Since we are interested only in qualitative patterns, rather than numeric quantities, we may ignore the denominator, since it is the same for all elements of the series.

In this case, the series produced by such a function would represent the contribution of each of successive information inputs, if there were more than one, to the changing of the knowledge structure. As we are fitting this within the context of Brookes' equation. 'contribution' does not mean information content measured in bits – that could be assessed separately – but rather the extent to which each input brings about the change. This allows for possibility that there may be several discrete information inputs bringing about the knowledge structure change. Indeed, Brookes specifically allowed for this [15, 8] in an early variant of his equation, showing successive inputs:

$I_1 + (S)_0 \wedge (S)_1$

$I_2 + (S)_1 \wedge (S)_2$

and so on.

If we knew the value of the exponent s – which we do not – we could directly calculate this contribution. Rather, we will consider what patterns emerge from varying the of values of s, using the simplest form of the series, with values $1 / k^s$, and noting that for negative exponents

$$a^{-m} = 1 / a^m$$

and that

$$a^0 = 1$$

The value of the exponent may range, in principle, from plus infinity to minus infinity. In order to see the overall pattern, in the simplest way, we take its value to be:

- large and positive, say 100
- small and positive, say 2
- zero
- small and negative, say -2
- large and negative, say 100

Writing out the first five terms of the series in each case, in factional exponential form, we can see the pattern without needing to calculate a numerical value.

k					
100	1	$1/2^{100}$	$1/3^{100}$	$1/4^{100}$	$1/5^{100}$
2	1	$1/2^2$	$1/3^2$	$1/4^2$	$1/5^2$
0	1	1	1	1	1
-2	1	2^2	3^2	4^2	5^2
-100	1	2^{100}	3^{100}	4^{100}	5^{100}

Positive, negative and zero exponents result in series for which the values respectively decrease, increase and remain constant, as the element number increases.

At high absolute values of k, the 'value' of the series is concentrated in the first and last elements respectively, while at low absolute values there is a smoother transition. For example, with a value of k=2, the series takes the values

1, 1 / 4, 1 / 9, 1 / 16 …..

corresponding to information inputs with values

1, 0.25, 0.11, 0.06 ….

This very simply application of a power law function within Brookes' equation suggests five main categories of information input, which are given the labels i1-i5:

i1 – large positive exponent

single information input, at start of series

i2 – small positive exponent

sequential decreasing information inputs

i3 – zero exponent

sequential equal information inputs

i4 – small negative exponent

sequential increasing information inputs

i5 – large negative exponent

single information input, at end of series

It should be emphasised here that the 'size' of the information input is not measured objectively in bits; that could be done, but is a separate issue. We are measuring size here, in the Brookes' equation context, by the cognitive impact on the recipient.

These five categories are not, of course, absolute. However, it is clear that only a small increase in the absolute value of the exponent results in a qualitative **i1** or **i5** categorisation. There is little scope for adding extra 'transitional' categories, which suggests that these five categories are a natural consequence of applying a power law function within Brookes' equation.

The series have been presented in the usual ordered form, but – thinking generally – there is no reason why the information input of series should be perceived in this way by the recipient. Any form of order is, in principle, possible; we will consider only the one which is most distinct from the ordered series, that in which the entries are randomised.

Randomising the entries of **i3** makes no difference to the pattern, since the entries are equal in magnitude.

Randomising the entries of **i2** or **i4** gives a new pattern; a series of information inputs with significant value arriving in random order.

Randomising the entries of **i1** or **i5** also gives a new pattern: a single significant information input, arriving randomly among inputs of no value.

We denote these as two new categories:

i6 – small exponent, series randomised

randomly arranged information inputs

i7 – large exponent, series randomised

single information input, randomly positioned

We now to consider both the validity and the necessity of these results. Do these seven categories match our knowledge of information access, and are these modes of access which they do not capture? And is it necessary to use this formalism to derive this categorisation?

Considering the categories in turn, the patterns of **i2** and **i4** are both characteristic of *search*. The distinction lies in the cognitive aspects. **i2**, with its decreasing inputs of information, may be regarded as a **confirmatory search**; the first input gives most of the information, with subsequent inputs adding correspondingly less. **i4** is best described as a **consolidating search**, perhaps in a area of knowledge new to the searcher; as more information inputs are added, they become of increasing value, as they can be integrated with the developing knowledge structure.

i1 can also be seen as a form of search, a particularly effective one, which provides the full information in the first input; it is typical of a **reference**, or **known-item**, search, satisfied by a single input.

The randomly arriving information inputs of **i6** seem characteristic of **browsing**, of the purposive or directed variety.

i7, with its single isolated randomly arriving information input denotes the kind of information access described as **accidental**, **serendipitous** or **encountering**.

I5, in which the single information input arrives at the end of a series of negligible inputs, is best thought of as the **eureka** situation, the earlier inputs having prepared the cognitive structure for the innovative leap.

Finally, the pattern of **i3** reflects a **monitoring** situation, with a series of inputs of equivalent value; most likely some routine or comparative access.

It seems, therefore, that all of these patterns can be equated to clearly understood patterns of information access. They encompass searching, in its precise and more general forms, browsing and accidental discovery, creative innovation, and routine information monitoring; a comprehensive coverage of information access from a cognitive perspective.

While, as noted by the references above, several of these patterns have been described in the literature of information behaviour and use, this comprehensive set of patterns has never been set out. This justifies the value of this extension of Brookes' equation.

It will be noted that this treatment does not include the situation where no information at all is acquired in any way. This is reasonable and appropriate, since the equation relates only to those situations where there **is** some information, having some effect.

3. ΔS: a cognitive map

The ΔS term in Brookes equation represents an internal, cognitive knowledge structure. As with our treatment of the ΔI term, we will make some minimal assumptions, and see where they lead.

Brookes was unspecific about how knowledge structures were to be represented, other than describing them generally as cognitive maps. A very common way of dealing with this issue in the information sciences has been to adopt the idea of a mental map or mental model, to describe an internal knowledge structure; a few examples typify this approach [16, 17, 18, 19, 20]. This is most commonly an arrangement, more or less complex according to need, of concepts joined by relationship links. The nature of the concepts and of the links may be described in varying degrees of detail, and the strength of, or confidence in, both concepts and links may or may not be assessed.

If we take the simplest possible form of this representation – and we have no *a priori* reason for doing otherwise - we can see that there are three general ways in which it may be changed, which we label, in a similar way to the information inputs, as **s1**, **s2** and **s3**.

s1 the pattern of concepts and links is unchanged, though the strength of any concept or link may be increased or decreased

s2 at least one concept or link is added, but none is removed; **s1** may also occur

s3 at least one concept or link is removed; **s1** and/or **s2** may also occur

s1, which we will call a **confidence** change, implies that the information input has increased or decreased the strength of some part of the knowledge structure. In including this possibility we follow Brookes [1] who argued that the change is knowledge structure may not be additive, but might involve a change of structure. Cole [6], in adopting Brookes equation as part of a operationalization of the concept of information, insisted that the change must be a change of structure, and that the change which we call **s1** is better regarded as 'conforming' input or data. In the interests of generality, we will stick with Brookes' idea.

s2, which we will call an **expansion** change, denotes an increase in the extent of the knowledge structure, with new concepts or links being added as a result of input information

s3, which we will call a **disruptive** change, denotes a qualitative change in the pattern of the knowledge structure.

It would be possible to give a more detailed or sophisticated categorisation of changes of knowledge structure, but any such would be dependent on the nature of some particular mental map; in any event they are not necessary for the argument here. Similarly, it would be possible to extend this treatment by giving some quantitative measure of the extent of the expansion or the structural change; again this is not necessary for our purposes. This simple, and all-inclusive, three-way categorisation will suffice.

Brookes, as he made clear, derived his equation very much in the context of a belief in Karl Popper's philosophy [1, 8]. It is therefore interesting to note that **s1**, **s2** and **s3** can be interpreted in Popperian terms - see, for example, O'Hear [21, chapter 3] and Corvi [22, chapter 5] - as a failed falsification, the creation of a trial solution, and a successful falsification respectively.

4. A categorisation of knowledge changes

We now have a seven-fold categorisation of information inputs, and a three-fold characterisation of changes in knowledge structure. Putting these together, as the equation invites us to, gives the possibility of an overall qualitative pattern, with twenty-one categories. Some are better-established than others, in terms of being described in the literature, but all are coherent and comprehensible. This list of categories should be a useful stimulus to assessing whether they have yet been observed; where they have not, this may be a stimulus either to investigating them, or perhaps to considering whether new forms of information system and service are necessary to support them.

The 21 categories are listed and described in Appendix 1. A few examples will suffice to give the 'flavour'.

Category 3 (i3-s1) denotes the sequential adding of a series of equivalent information inputs, with the result that the existing knowledge structure is strengthened. This can be understood as a short-term event, for example a search for essentially the same information in different sources, or longer-term as a routine monitoring of information over time, in either case confirming an existing cognitive viewpoint.

Category 8 (i1-s2) denotes an immediate single information input, which causes an expansion of the knowledge structure. This can be understood as a targeted search, or reference /known-item enquiry, perhaps for a specific piece of factual information, which is added into an established structure.

Category 11 (i4-s2**)** denotes a sequential adding of search results, each of greater significance, with the result of an expanded knowledge structure. This would cover the acquisition of information from scratch in a new area, where each added element had greater significance as it was fitted within the growing knowledge structure.

Category 19 (i5-s3) denotes the 'Eureka' moment, when a sequence of information inputs appear to have no effect, until the final one causes a disruptive change in the knowledge structure. Perhaps we should call it enlightenment.

The example of category 3 above shows that these categories of knowledge changes may each represent the result of different activities, tasks and time scales. Furthermore, it may not be easy to determine, solely from a study of information-related behaviour, what category is relevant.

For example, Nicholas, Huntington, Williams and Dobrowolski [23] identified a behaviour denoted as 'checking'; repeatedly examining what was apparently the same information on different web sites. It is tempting, and probably correct in many cases, to equate this to Category 2 (i2-s1) : a series of information inputs of decreasing significance –since all that is being found is repetition of the first information leading to confirmation of an established structure. However, it may be that the very fact of the wide repetition of this piece of information is of interest for some reason; in that case, the change might be expansion, and the sequentially added information items might be of equal or increasing significance, so that the knowledge change might be category 10 or 11. This is an indication of the power of this categorisation as a tool for examining just how well we understand information behaviour.

It is also true to say that it may possible to identify with confidence the nature of the information inputs, but remain in ignorance of the nature of the knowledge structure change. An example of this is the identification by Erdalez [24] of the phenomenon of 'encountering', the serendipitous acquisition of information. This seems to equate to information change i7, but its consequences in terms of structure change must remain unknown without knowledge of the particular circumstances; the result could be any of s1, s2 or s3.

5. Conclusions

By making the assumptions that information inputs can be represented by the simplest form of power law, and knowledge structures by the simplest form of mental map, and applying these within the framework of Brookes equation, a qualitative characterisation of the processes represented by the equation can be derived. This has a pattern of seven forms of information input and three forms of change in knowledge structure, giving twenty-one categories overall. It would be possible to choose more complex alternatives to represent both function and structure, but there is no rationale for doing so, and it would result in a categorisation which was more complicated and multifaceted, to the point of being unhelpful.

These may be used directly, as an explanatory tool and taxonomy for studying and summarising information behaviour and use, and – in particular – for showing when understanding is less than full. Furthermore, since they result from a very simplistic semi-quantisation of Brookes' equation, they offer the possibility of being developed into more sophisticated and quantitative treatments.

6. References

[1] B.C. Brookes, The foundations of information science, Part 1. Philosophical aspects, *Journal of Information Science* 2(3/4) (1980) 125-134

[2] N.J. Belkin, The cognitive viewpoint in information science, *Journal of Information Science*, 16(1) (1990) 11-15

[3] P. Ingwersen, Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory, *Journal of Documentation* 52(1) (1996) 3-50

[4] N. Ford, Modeling cognitive processes in information seeking: from Popper to Pask, *Journal of the American Society for Information Science and Technology*, 55(9) (2004) 769-782

[5] P. Ingwersen and K. Järvelin, *The Turn: Integration of information seeking and retrieval in context* [Springer, Dordrecht, 2005)

[6] C. Cole, Operationalizing the notion of information as a subjective construct, *Journal of the American Society for Information Science*, 45(7) (1994) 465-476

[7] P. Ingwersen, *Information retrieval interaction*, (Taylor Graham, London, 1992)

[8] R.J. Todd, Back to our beginnings: information utilization, Bertram Brookes and the fundamental equation of information science, *Information Processing and Management*, 35(6) (1994) 851-870

[9] I. Cornelius, Theorizing information for information science, *Annual Review of Information Science and Technology*, 36 (2002) 393-425

[10] C. Cole, Calculating the information content of an information process for a domain expert using Shannon's mathematical theory of communication: a preliminary analysis, *Information Processing and Management*, 33(6) (1997) 715-726

[11] S. Neill, Brookes, Popper and objective knowledge, *Journal of Information Science*, 4(1) (1982) 33-39

[12] R.J. Todd. Utilization of heroin information by adolescent girls in Australia: a cognitive analysis, *Journal of the American Society of Information Science*, 50(1) (1999) 10-23

[13] D. Bawden, Smoother pebbles and the shoulders of giants: the developing foundations of information science, *Journal of Information Science*, 34(4) (2008) 415-426

[14] L. Egghe, *Power Laws in the information production process: Lotkaian informatics* (Elsevier, Amsterdam, 2005)

[15] B. Brookes, Robert Fairthorne and the scope of information science, *Journal of Documentation*, 30(2) (1974) 139-152

[16] L. Westbrook, Mental models: a theoretical overview and preliminary study, *Journal of Information Science*, 32(6) (2006) 563-579

[17] S. Makri, A. Blandford, J. Gow, J. Rimmer, C. Warwick and G. Buchanan, A library or just another information resource? A case study of users' mental models of traditional and digital libraries, *Journal of the American Society for Information Science and Technology*, 58(3) (2007) 433-445

[18] C. Cole, Y. Lin, J. Leide, A. Large and J. Beheshti, A classification of mental models of undergraduates seeking information for a course essay in history and psychology: preliminary investigation into aligning their mental models with online thesauri, *Journal of the American Society for Information Science and Technology*, 58(13) (2007) 2092-2104

[19] W. He, S. Erdelez, F. Wang and C. Shyu, The effects of conceptual description and search practice on users' mental models and information seeking in a case-based reasoning system, *Information Processing and Management*, 44(1) (2008) 294-309

[20] J. Pisanski and M. Žumer, Mental models of the bibliographic universe. Part 1: mental models of descriptions, *Journal of Documentation*, 66(5) (2010) 643-667

[21] A. O'Hear, *Karl Popper* (Routledge, London, 1980)

[22] R. Corvi, *An introduction to the thought of Karl Popper*, translated by P. Camiller, (Routledge, London, 1997)

[23] D. Nicholas, P. Huntington, P. Williams and T. Dobrowolski, Re-appraising information seeking behaviour in a digital environment – bouncers, checkers, returnees and the like, *Journal of Documentation*, 60(1) (2004) 24-43

[24] S. Erdelez, Investigation of information encountering in the controlled research environment, *Information Processing and Management*, 40(6) (2004) 1013-1025

Appendix 1

Categorisation of knowledge changes

1 **i1 – s1**
single information input at start, with confidence change

2 **i2 – s1**
sequential decreasing information inputs, with confidence change

3 **i3 – s1**
sequential equal information inputs, with confidence change

4 **i4 – s1**
sequential increasing information inputs, with confidence change

5 **i5 – s1**
single information input, at end of series, with confidence change

6 **i6 – s1**
randomly arranged information inputs, with confidence change

7 **i7 – s1**
single information input, randomly positioned, with confidence change

8 **i1 – s2**
single information input, at start of series, with expansion change

9 **i2 – s2**
sequential decreasing information inputs, with expansion change

10 **i3 – s2**
sequential equal information inputs, with expansion change

11 **i4 – s2**
sequential increasing information inputs, with expansion change

12 **i5 – s2**
single information input, at end of series, with expansion change

13 **i6 – s2**
randomly arranged information inputs, with expansion change

14 **i7 – s2**
single information input, randomly positioned, with expansion change

15 **i1 – s3**
single information input, at start of series, with disruptive change

16 **i2 – s3**
sequential decreasing information inputs, with disruptive change

17 **i3 – s3**

sequential equal information inputs, with disruptive change

18 **i4 – s3**

sequential increasing information inputs, with disruptive change

19 **i5 – s3**

single information input, at end of series, with disruptive change

20 **i6 – s3**

randomly arranged information inputs, with disruptive change

21 **i7 – s3**

single information input, randomly positioned, with disruptive change