Simulation, and simulation experiments

Michael D. Heine

10.1 Introduction

The term 'simulation' is variously defined:

> 'Some simulations seem to be searches for structure, for important variables, for hypotheses. Others are tests of hypotheses. Still others trace the implications of a theory or simply enumerate the consequences of particular mechanisms, policies, procedures, or models.' (Chapman[1], p. 482)

As the terms 'system' and 'model'—even 'information'—which tend to occur with 'simulation' are also variously defined, the difficulties in communication are compounded. Since we have yet to agree to a system of thought that encompasses all human cognition, a degree of ambiguity is perhaps inevitable, given the very abstract connotation that simulation can have. A brief classification of meanings is as follows.

To some authors the term 'simulation' is associated with an essentially *verbal* analysis of an area of phenomena of interest—which is termed 'the system'. Or the 'system' may be a *sub-area* of the area of interest, delimitation of the sub-area being seen as an intrinsic part of the analysis. (The verbal description may be accompanied by diagrams representing, in an indicative rather than a rigorous way, the thoughts in the analysis.) To other authors, simulation is a branch of applied mathematics (or is *all* of applied mathematics), the concern being to represent real entities and relationships (i.e. observables) by numeric variables and analytical functions. Here simulation is an embodiment of the conventional approach of the physicist or engineer, and the definition is little different from that of classical science itself. 'Objectivity' in analysis is recognized—indeed is required to be so. The critical thing to yet other authors is that the techniques used to explore mathematical relationships are implemented by computers, often involving exploration of the cumulative effects of manipulating values taken randomly from a (0,1)-uniform continuous distribution. Here no particular concern is shown for either the problem of system delimitation or the aims of the study.

The apparent vagueness in 'the simulation approach' just described is a little illusory, for beneath the various definitions can be discerned general features of just one concept, both justifying our provisional use of the one term and encouraging us to relate it to information systems. First, the notion of 'optimization' provides one general feature. One does not vary either the representatives of the components of a system or the sequence of operations in a system, capriciously. One does so in order to identify a system that is 'best' according to some criterion. In that a notion of social utility is so implied, simulation could be seen as differing from the classical sciences—which are simply predictive systems of thought—and positioned rather more closely to technology. Yet it differs from the latter too in that 'judgement' is always a component of simulation in practice—a second general feature. This consists, at the least, in the specification of a criterion or of criteria by which optimization is to be judged; and will also consist in the delimitation of the 'system' to be studied and the identification of the system's essential components. If simulation could both *delimit* an area for analysis as well as *be* the analysis, this would involve attributing an 'intelligence' to it. Obviously it is the presence of human judgement together with some rationale for analysing phenomena that embodies this intelligence. (It would be 'an error of judgement' to omit to identify a variable that sensitively affected another variable by which optimality was to be judged.) This raises intriguing questions where the *subject* of the simulation is human decision-making, for example 'relevance-judging' or 'query-forming' in information retrieval. Here one is (in simulating same) making judgements about judgements. And since a formulation of a query (say) is in part a judgement about how an indexer will have chosen index terms, i.e. again a judgement about a judgement, we have a third tier of judgement. It seems that a recursive definition of simulation is needed to clarify things here, but to the writer's knowledge this has not yet been attempted. A third general feature of simulation is the admission, into the system description, of the notions of 'input' and 'output'. These entities, usually seen as separated in time, may be real artefacts (e.g. documents, money) or information (a message on a VDU) or decisions ('Get file XYZ'), to choose information retrieval instances. Lastly, and really in continuation of the first feature above, we note that one's interest in simulating systems is usually prompted by the possibility of *intervention* in the system: or even total control of it. Without an interventionist ethos we would be back in the realm of classical theoretical science.

In the face of this complexity, partly of the making of those who have written on simulation, the temptation is to wield an Occam's razor in the form of just one clear prescriptive definition of 'simulation'. In the writer's view however this would be a mistake in the particular context in which information retrieval experimentation finds itself: a context which is commonly denoted 'information science' but which might better be called 'the philosophy of information'. For the ambiguities of simulation in this context are indicative of the ambiguities that we need to overcome if information science is to develop. The characterization of 'information', 'variable', 'representation', 'system' as part of the human cognitive apparatus, and the attaching of abstract meanings to concepts such as 'observable' or 'objective viewpoint' or 'explaining power' or 'language' are just the kinds of things we need to do in order to develop information science. So that taking the *broadest* definition of information science we could say, more rhetorically, that the philosophical problem of saying what is meant by simulation, is the problem of saying what is meant by information science. (Indeed, one might develop a philosophy of *science* along these lines, by regarding simulation of the World as the function of science. This raises intriguing questions as to the proper place of information science. Possibly we could retain our primitive objectivity as the only objectivity that matters, and assert that one cannot make an objective study (i.e. simulation) of that objectivity. Science might then be construed as the simulation of the totality of information structures with which the World confronts us; and information science, as one species of that science, could sensibly be defined as the simulation of *the interactions of information structures*. Alternatively, we could assert that one can be objective about the objectivity that science assumes, and that objectivity about such objectivity constitutes information science: information is then a metascience of the character of 'simulation about simulation'. The issues are not just semantic: there is also the related question of whether obtaining a complete World picture is or is not a paradoxical requirement.) It would, of course, be perfectly feasible—and much simpler—to adopt an alternative stance, in which simulation was treated in just the narrow way, and where the simulation of information flow processes was treated as a branch of engineering—in much the same category as the simulation of fluttering aircraft wings, or of cash flow in an economy, say. But this would, the writer suggests, be both to ignore the opportunity for cross-fertilization between an enriched information science and our retrieval technology, and to ignore the arbitrariness entailed in singling out information retrieval experimentation for investigation. It would, in particular, beg the question that the philosophy of information to which information retrieval experimentation is capable of contributing, can contribute nothing in return.

It is suggested therefore that there is some virtue in leaving the concept of simulation, in the context of information science, as an intuitive one—at least for the time being. The hope is that general descriptions of and technology relevant to information transfer will develop from piecemeal simulation studies of all kinds in this area, with appropriate concepts and terminology crystallizing out from this.

To try to obtain a perspective on the kind of system that could be recognized for the purpose of simulation studies in the area of information retrieval we could write down the following:

System 1: The identifying, by a person, of an information need.

System 2: The expressing of a need by means of a verbal artefact (i.e. a 'question').

System 3: The recognizing of (i.e. perceiving of) 'relevance' between a given document record (e.g. an author + title, or author + abstract) and an information need.

System 4: The identifying of document attributes by an individual (the indexer) in relation to a document, that anticipate the verbal artefacts to be yielded in System 2, given that the document record would also be recognized as relevant by that person in System 3.

System 5: The matching of document records, as attributed under System 4, with verbal artefacts yielded by System 2, and the selective output of such records.

System 6: The preparing of documents by their authors, and the disseminating of documents.
System 7: The transferring of messages by means other than documents.
System 8: The transferring of messages (with Systems 6 and 7 as subsystems).

Some of the above can, individually, be broken down into subsystems: System 8 obviously so, and System 5 into (for example) a purely record-manipulative subsystem, a *logical* subsystem, and an economic subsystem. Conversely, more general systems, of which the above could be regarded as components, can be defined. For example, combining Systems 2 and 4 provides one challenging problem here in that two parties: the indexer and the searcher, are attempting to predict the behaviour of the other, i.e. a 'gaming' system is implied. One side of this has recently been treated by Maron and W. S. Cooper (and also treated by writers on automatic indexing). Cooper[2] for example has seen indexing as a 'thought-experiment' by the indexer, in which the query terms to be used by the enquirer are anticipated. The other side, the anticipation by the enquirer of indexing terms that will have been used by the indexer (often over a lengthy period of time) is implicit in work on query optimization—for example that of Ide, Rocchio and Salton reviewed by Salton[3] and that of Barker, Veal and Wyatt[4] and Vernimb[5]. Identification of the latter system (System 4) and the composite system has, the writer suggested, been inhibited by the continued use of the misleading phrase 'relevance to a question' which has both obscured the concept of the question as a *variable*, and implied that relevance judgements are capable of being based, unambiguously, on a purely verbal construct[6]. An ideal would be the recognition of a generalized system taking in all of the above systems, as has been attempted in various semi-intuitive representations by some authors, e.g. Vickery[7]. Such generalized systems are usually represented as 'circular' in form, in that the output of the system (documents) contributes to a corpus of recorded knowledge, in relationship to which new information needs are recognized.

It is suggested that two consequences of the above discussion are: (1) Judgements as to the informational components and decision-making components of simulation take on a peculiar significance in information science, where such components are in fact the main conceptual targets of the science itself. To attempt to 'simulate' relationships between such components introduces almost a paradoxical situation (we are studying by systems means what it is to be a system). The effect of this, it is suggested, is that we should avoid strict definitions of simulation, and be aware that resolution of the difficulties may provide us with the conceptual roots that we seek. (2) Information retrieval experiments in the conventional sense—the Cranfield sense—involving study of the analyses of the effect on retrieval performance of altering the database (e.g. the depth of indexing in records), or say the boolean expressions representing users' information needs, relate to a subsystem of a larger system of information flow—using the term 'information' intuitively. This subsystem, the subsystem of 'retrieval from a database', appears to be an amalgam of what we have labelled as Systems 2, 3, 4 and 5, and could be labelled System 9 say.

We now contrast simulation in the narrower, mathematical sense, with investigation and experimentation, and briefly comment on the term 'simulation experiment'. All are processes that we recognize to be mixed 'cognitive/behavioural'. They all help us 'get on better' with the World in which we find ourselves by acquiring information for us, i.e. they alter the mimetic structures that govern our individual and social responses to the constraints and opportunities offered by other systems. However, investigation and experiment (treating these as similar though distinct processes) have two features that simulation does not: an experimental apparatus is needed for them to be implemented (even though, for investigation of document/user interaction, say, the apparatus may just be a file of records), and no suppositions are made as to how the information acquired is *generated* (one can for example, experimentally measure the acceleration 'due to gravity' without knowing how the system of primitive entities determining the acceleration are affecting the apparatus and so determining the data). Simulation, on the other hand, does not require an apparatus, and does concern itself with how information (data) is generated by a system. Suppositions are made about the entities making up the system (though the entities are regarded as 'constructions' rather than 'descriptions', so that suppositions is not quite the right term), about the relationships between entities, and about the effects of system input upon entities and (possibly) relationships. This definitional structure is then used to predict the output or outputs of the system, which can be described as information or data. So that unlike experiment and investigation, the determining entities in simulation work are not treated as primitive ones but as objects of study. That is essentially the strength and weakness of simulation work—as it is of all 'theoretical' study: the objects of interest and experimental study are made explicit but they remain constructs. This may seem a trivial or fine point, since in practice when simulation is applied at the human level (queues for tickets, say) or in an area of technology where 'laws' are well established, all entities of relevance seem clearly evident, and some of them are under our control (e.g. the number of serving booths) or can at least be influenced by us. But in view of the unclear foundations of information science, it seems essential to emphasize that the outcomes of simulation work, since they are based on a human construct, can never 'surprise' us (and so inform us) as much as experimental results can. We can be a *bit* surprised by the results of a simulation study (e.g. in regard to a pattern of symmetry or an instance of invariance that we missed in an experiment) but never *very* surprised, since the simulation explores a structure that we ourselves created: the results are, in that sense, tautological or necessary. Just as mathematics as an edifice of thought is inviolate and 'safe' (that's what it's there for), so is a simulation study. Both lack the open (i.e. receptive to amendment) syntax of science, a syntax which encourages information feedback that modifies and invigorates its structure when that information is obtained from experimentation.

Returning to terminology now, we define a simulation study as a 'simulation experiment' when the system's components (e.g. the parameters of probability distributions) are given certain values, or are explored in a certain order, and the consequences of same are noted. It is, in regard to Cranfield-type experimentation in information retrieval, a moot point whether some of such work should be described as 'experiment' or 'simulation', simply because it is not natural phenomena but man-created phenomena that are being explored. If an information retrieval experiment examines the effect of variation of relevance judgements (for an algorithmically-defined query say, and a set of information needs) on Recall at the Precision value 0.60, this might properly be termed an 'experiment' since the relevance-judgements, one's object of study, are not under the control of the experimenter. But if, for a given exercise of relevance-judgement by a person (applied to all items in a test-collection, say) we examined the relative effects on Recall (at $P=0.60$) of varying the document weighting expression, this might better be termed a 'simulation experiment'. In the latter instance, all components of the system are (a) known, and (b) under control. The outcome of the investigation is determined by logic: it is 'pre-determined'. (We would probably use a computer to obtain the output data though that is more incidental, and that alone would not justify a process being termed a simulation.)

In this introduction, the reasons or motivations for undertaking simulation studies (in the narrower senses) of information retrieval, in preference to experimental studies, have not been discussed in detail. This is because they are fully discussed in general terms in the standard texts on simulation (e.g. Churchman[8], Martin[9], Gordon[10]), and because the reasons tend to carry little conviction until one has actually undertaken a simulation study—at least in the author's experience. But briefly it might be claimed that (a) the simulation study itself requires that a formal representation of the system (called a 'model' by some writers) be arrived at, this itself giving valuable insight into the system, (b) manipulation of data and paths within a conceptual framework of a system (i.e. within a formal structure or model) is more economical of effort, money and time than manipulation of the real system so represented and measurement of data pertaining to it (i.e. than experimentation on the system)—which of course begs the question of the validity of the formal structure, and (c) the simulation suggests new areas for observational (experimental and investigative) work predicated on the validity of the constructions that it comprises.

In summary, simulation in its broadest sense is of interest to information retrieval workers because of the very uncertainty of its definition; because it provokes our interest in the conceptual roots of representation and transfer of information (or should we say, of representation and the transfer of representation)—the main topics of what we call 'information science'. This is of concern in information retrieval experiments both because such experiments appear to serve as a prototype for experiments on information transfer construed more generally, and because more formal (theoretical) study of information transfer may give insights into the process of information retrieval as we usually regard it. Our experiments and investigations of document transfer at the macroscopic 'human' level are into instances of information transfer that are perhaps artificially circumscribed. On the other hand, simulation in its narrowest sense, that of the representation of randomness in relationships between people, documents, document attributes, and logical expressions representing 'information needs', is of interest and value in challenging the terms and concepts we use, in distinguishing between tautological findings (i.e. findings simply a consequence of the language of description) and findings that are not tautological ('scientific' findings), and in suggesting new experiments consistent with conjectures within the language of description. But truths suggested by a simulation experiment are always suspect—in that the structure of the simulation is one that we created—and if important should be tested by experiment, the arbiter of truth.

10.2 Examples of simulation models in information retrieval studies

In order to demonstrate the variety of the 'style' of the simulation approach in information retrieval we describe three examples of simulation models relating to it. Except for the first example, no details are given as to how the models can be implemented on a computer, i.e. expressed as a sequence of instructions. The first example serves to show how a widespread program package, the Statistical Package for the Social Sciences (SPSS) can be used for simple simulation purposes. The second example is a paraphrase of the treatment of Morse's browsing model by Salton (Morse[11], Salton[12]). The third example represents a novel extension of the model put forward by Swets[13,14] interpreted in a discrete formalism. The examples relate to three very different areas in information retrieval: the speed with which documents are supplied from a library network (through some given library); the enhancement of 'browsing' in a collection, achieved by relegating little-used material from it; and the distribution of the effectiveness (expressed as a pair of Recall–Precision values) of boolean search expressions input to a database, when the terms of which the search expression is made up are given.

Example 1 (use of SPSS as a simple simulation tool)

Orr has suggested the possibility of systematically measuring the speed of supply of documents through a given library local to the user, where the library is (as is usual) connected to one or more other libraries which can supply documents not available locally[15]. Each item in a sample of documents, allegedly a random sample of documents needed by the clients of a given library, is assigned a 'delivery time'. This is the time taken to supply the item—whether from a library local to the user or from a 'connected' library. The delivery time is in fact a label for an interval into which the actual time taken is placed, the intervals being approximately $(10^{n-1}, 10^{n})$ minutes, $n=1,2,3,4,5$. (It is considered that these intervals correspond reasonably closely to our *subjective* notions of document delivery time, which a straight arithmetic scale does not.) Orr's approach is especially interesting in that (a) it explicitly treats document delivery *time* as an indicator of library effectiveness, (b) it gives a measure of overall effectiveness unbiased by an existing pattern of demand (as distinct from need), and (c) it measures not the effectiveness of a library 'in isolation' but its effectiveness *contingent on* the strength of its connections to other document sources and the extent of those sources. The difficulties in applying the method appear to be, principally, those of identifying a convincing sample design strategy, and accommodating substitutibility of information demand into the method.

Denote document delivery time by *TG* (so that the possible values of *TG* are 1, 2, 3, 4 and 5), and define a new variable as 125-25*TG*. The mean value of the new variable is known as the 'Capability Index' of the library (contingent on a specified backup system), as defined by Orr, and is denoted by *CI*, i.e. $CI = E(125\text{-}25TG)$. Thus when delivery times all have the value 5 (the least effective library possible), *CI* has the value 0, and when delivery times are the shortest possible, *CI* has the value 100. A large sample of requests for documents is examined and data obtained on the delivery time for each document, for some given library connected to the document supply system. Details such as attributes of the document and the requestor are also noted. It is required to know how various policy-changes affecting the system are likely to influence its *CI*-value, as part of a policy review of the system. To simulate the policy changes we can add a random number (strictly add the value of a random variable *U*, uniform in the interval (0,1)) to each set of numeric or qualitative values characterizing each document. The data set would then comprise (say) the values of variables describing the user, the document, a value for the delivery time, and the random number. (To do this, a file of SPSS data could be read by a simple FORTRAN or ALGOL program accessing a suitable program package, such as that of the Numerical Algorithms Group (NAG), outputting the enriched data to a new file.) The 'IF' command of SPSS can then be used to reassign the value of *TG* on the basis of (a) the user and document data for the case concerned (as appropriate to the policy change of interest), (b) the value of *U* recorded for the case, and (c) a specified threshold value arrived at by examination of independent evidence. The variable $XCI = 125\text{-}25{*}TG$, is computed *after* this reassignment of *TG* (using the 'COMPUTE' command). Lastly, the 'STATISTICS' command of SPSS will yield a value for the mean value of *XCI*, which happens to be the value for Capability Index, *CI*, that we seek. For example suppose the policy option being considered is 'obtain all requests for documents that are papers in serials and are not held by the library, and which are requested by users of status S, as follows: (1) as photocopies, and (2) from the interlending source J' and suppose that the independently obtained evidence is that in such cases 65 per cent of such requests are delivered in time $TG = 4$, the rest in time $TG = 5$. Then we would test for the appropriateness of each document in the sample to this policy option by using an IF statement to identify documents that were both photocopies of serial papers and requested by S-type users and, in addition, for which $U < 0.65$ was true. In those cases we would reassign to *TG* the value 4. If the document were a photocopy of a serial paper and requested by an S-type user but $U < 0.65$ was false, *TG* would be given the value 5. After these reassignments, the value of *CI* would be calculated as usual, the new value of it indicating the likely effect of implementing the new policy option when *all* needs are considered. Examples of other policy options that might be considered are those of giving users direct access to other systems,extending the hours of opening of the local collection and extending the loan period for locally-owned documents. We note that the use of SPSS in this way assumes independence between certain random variables implied by the raw data—a reasonable assumption if no contrary evidence is available.

Example 2 (Morse's model of browsing in relegating collections, as treated by Salton)

This example offers a description of a system (document supply system, or document record supply system) which has the following properties: (1) the system follows a policy of relegating items from it when the rate of usage of such items falls below some threshold value, and (2) usage is of a 'browsing' nature. By 'browsing' we mean a process whereby the user identifies relevant documents by examining documents (or records) chosen randomly (uniformly) from the collection; at least the simulation *represents* what we term browsing in this manner. (Relegation of documents could in practice be undertaken using an 'age of document' criterion, say, if it were found that relevant items (i.e. items used) were found to have a mean age differing significantly from the mean age of non-relevant items.) Browsing usage so defined is not distinguished from usage of other kinds—which of course constitutes a severe limitation of the model. The collection that 'remains' after less-useful documents have been relegated is referred to as a 'reconcentrated' collection. Our interest is in the *enhancement to browsing* achieved by such relegation.

It was shown by Morse that the probability of identifying a relevant item, placed at random in a collection or database, is:

$$P = 1 - \exp(-\delta t/n)$$

where t is the time taken, δ is the search rate, and N is the size of the database, and the enquirer searches randomly. (The analogy of most interest in information retrieval work in its narrower sense is perhaps where the relevant items are placed at random in a retrieved set of records, the size of this being proportional to the size of the parent file, it might be assumed.) Suppose that the collection or database is divided into a more-relevant section and a less-relevant section much as the MEDLARS database is divided into the MEDLINE file and a set of BACKFILEs; and denote the estimated mean numbers of items of relevance in the whole collection and the reconcentrated collection by $\bar{E}$ and $\bar{E}_r$. Also denote the mean numbers of relevant items *identified* by searching the whole collection and the reconcentrated collection for a time t by $\bar{S}(t)$ and $\bar{S}_r(t)$. Then:

$$\bar{S}(t) = \bar{E}(1 - \exp(-\delta t/N));$$

and

$$\bar{S}_r(t) = \bar{E}_r(1 - \exp(-\delta t/xN))$$

where xN is the size of the reconcentrated collection. Using simplifying assumptions it can then be deduced that the relationship between $\bar{E}$, $\bar{E}_r$ and x is:

$$\bar{E}_r = x\bar{E}(1 + \ln(1/x))$$

The effectiveness of the subdivided collection, i.e. the effectiveness of choosing a value for x (other things being equal) may then be interpreted as either (a) the ratio $\bar{S}_s(t)/\bar{S}(t)$ (the ratio of the numbers of items obtained in a given time from the reconcentrated collection and the main collection), or (b) the ratio of the Recall values, $\bar{E}_r/\bar{E}$ (which is necessarily less than 1).

This type of model appears to be appropriate to both the problem of optimum online file size, and the problem of optimum local library size: each is an analogue of the other. The latter problem (discussed for example by Gore[16], or the United Kingdom University Grants Committee[17]) has received particular attention in recent years in the area of academic librarianship with the enforced abandonment, for economic reasons, of the goal of local self-sufficiency in libraries in favour of both (a) systematic stock relegation and (b) heavier reliance on both local 'closed access' collections and more remote regional and national collections.

Example 3 (The 'logical surface' and 'document weighting surface' of a set of terms)

The essence of information retrieval is that a person, in recognition of an information need, perceives a set of document attributes that in his judgement best distinguishes the documents relevant to that need from other documents. Suppose our interest is in the overall sensitivity of the Recall–Precision outcome of retrieval to (1) the choice of logical expression embodying a given set of N attributes (e.g. terms), and (2) the choice of (a) document weighting function, and (b) threshold value, for a set of N terms. We shall approach the problems using the system representation of Swets, as extended by the author[18], and note that the general equivalence between the two retrieval strategies just mentioned was first pointed out by Angione[19]. We note also that the problems identified are subproblems of broader problems in which the *identity* of the set of N terms is allowed to vary.

Suppose, for illustration, we take the value of N to be 4. (So subsequent expressions such as 2^4 can be generalized to 2^N.) The four terms comprising the query are denoted by T1, T2, T3 and T4; the query itself, construed as such as set, by Q. The distinct elementary logical conjuncts of these terms are 2^4 in number, examples of same being:

$$\mathrm{T1} \wedge \mathrm{T2} \wedge \mathrm{T3} \wedge \mathrm{T4} \quad \text{or} \quad \mathrm{T1} \wedge \neg\mathrm{T2} \wedge \mathrm{T3} \wedge \neg\mathrm{T4}$$

Here, '$\wedge$' denotes conjunction and '$\neg$' denotes negation. These 16 elementary conjuncts may be disjoined (i.e. ORed) together in any combination. Accordingly, for 4 search terms there are exactly ${}^{16}C_0 + {}^{16}C_1 + {}^{16}C_2 + \cdots + {}^{16}C_{16}$, or $2^{16} = 65536$ *distinct* boolean expressions with which an enquirer may probe a database. The elementary conjunct in which all search terms are negated, namely $\neg\mathrm{T1} \wedge \neg\mathrm{T2} \wedge \neg\mathrm{T3} \wedge \neg\mathrm{T4}$, is unlikely to be employed in practice, so that this total might reasonably be modified to $2^{(2^4-1)}$. Rejection of particular disjunctions of the elementary conjuncts other than the all-negated one might also be reasonable, but experimentally obtained evidence of user behaviour in this regard would be needed to justify particular choices. In the absence of such evidence it seems reasonable to proceed without arbitrary rejection of any of the possible boolean expressions that might be used—other than those involving the all-negated elementary conjunct. Our interest is in the probability distribution that this set of search expressions defines over the Recall–Precision 'area' (i.e. over the area $(0,1) \times (0,1)$). This will be the distribution on Recall–Precision outcomes when the *form* of the boolean expression is chosen *arbitrarily* by the enquirer, but the *component terms* of the expression are *fixed*. The surface will in general be specific to a given instance of information need (defined objectively as a partitioning of the data base) and a given query set Q.

Before taking the above further, we examine the second problem. A document weighting function (DWF) acts so as to order, or partially order, the elementary conjuncts that we have described. This is so since a DWF serves to map the values of $Q \cap T_d$, where T_d denotes the set of terms attached to a document, to the real numbers. These values are each identifiable with a set of elementary conjuncts of the terms that make up Q. This mapping could be done by, for example, summing the non-negated terms that are attached in the document—the Cranfield 'co-ordination level' function—or by forming the product of, or summing, the specificities of the non-negated terms attached to the document. If the DWF is such that more than one elementary conjunct maps to the *same* numeric value (which is certainly true for the co-ordination level function) the ordering of elementary conjuncts is a partial ordering. The outcome values of DWFs are of course the 'document weights' of interest, each document being mapped to exactly one value. (The concept of 'term weight' is a superfluous one and in the author's view better avoided, since it obscures the essential relationship of interest: the association of each *document* with one or more elementary conjuncts.)

Our interest is thus in relating each of the retrieval strategies to values of Recall and Precision as the latter is determined by identification of selected elementary conjuncts of query terms. The ways in which this can be done become apparent if we specify all the elementary conjuncts and associate each such conjunct with a pair of probabilities: the probability that a relevant document is mapped by the function $Q \cap T_d$ to that conjunct, and the probability that a non-relevant document is mapped to that conjunct. (We should, strictly, refer to two functions defined by the *expression* $Q \cap T_d$, since the domain differs in the two cases.) For $N=4$ the list is as in *Table 10.1* where $\sum r_i = 1 = \sum f_i$. The Swetsian probability distributions are defined once each conjunct has been mapped to a real number by a DWF, the numbers not necessarily being distinct. (This is, at least, one interpretation of Swet's contribution. In the experimental results re-analysed by Swets, specific data sets such as the above were in fact confounded before the distributions were identified.) These distributions, as a pair, define a set of Recall–Precision co-ordinates, i.e. a graph, if a threshold value is allowed to explore the various outcome values and a document is retrieved only when its weight exceeds that threshold value. For a threshold value of T, the Recall and Fallout values are:

$$R = \sum_{i=T}^{16} r_i \quad \text{and} \quad F = \sum_{i=T}^{16} f_i$$

with the Precision value following from $P = G/[G + (1-G)(F/R)]$. Such graphs are defined by a particular instance of information need (evidenced through a partitioned database), a particular query (defined as a set of terms), and a DWF. Since the elementary conjuncts can be ranked (permuted) in $(2^4)!$ ways, there are $(2^4)!$ subsets of DWFs distinguishable through the rankings of elementary conjuncts that they effect, of those DWFs that yield 2^4 distinct outcome values. DWFs can also be classified by the Recall–Precision graphs that they effect, for a particular need and query. Each graph possesses at most 2^4 distinct R–P co-ordinates. (One says 'at most' since if '(0,0)' values of (r_i, f_i) are present, the cumulative values of R and F will be unaffected by incrementation of the threshold past such pairs and so the R–P graph will lose one co-ordinate for each such pair. As N increases, the number of such probability pairs will increase.) The number of distinct R–P graphs is at most $(2^4)!$ Possibly this is the exact number, but since Precision depends only on the *ratio* of Fallout to Recall, one suspects that some Recall–Precision graphs may fortuitously be identical. (Uncertainties of this nature are in fact an argument for using simulation techniques: both to bypass establishment of their truths by formal means, and to try to establish exceptions to general statements.)

TABLE 10.1

Rank value	*Elementary conjunct*	*Probability for set of relevant documents*	*Probability for set of non-relevant documents*
a	T1 ∧ T2 ∧ T3 ∧ T4	r_a	f_a
b	T1 ∧ T2 ∧ T3 ∧ ¬T4	r_b	f_b
c	T1 ∧ T2 ∧ ¬T3 ∧ T4	r_c	f_c
d	T1 ∧ T2 ∧ ¬T3 ∧ ¬T4	r_d	f_d
e	T1 ∧ ¬T2 ∧ T3 ∧ T4	r_e	f_e
f	T1 ∧ ¬T2 ∧ T3 ∧ ¬T4	r_f	f_f
g	T1 ∧ ¬T2 ∧ ¬T3 ∧ T4	r_g	f_g
h	T1 ∧ ¬T2 ∧ ¬T3 ∧ ¬T4	r_h	f_h
i	¬T1 ∧ T2 ∧ T3 ∧ T4	r_i	f_i
j	¬T1 ∧ T2 ∧ T3 ∧ ¬T4	r_j	f_j
k	¬T1 ∧ T2 ∧ ¬T3 ∧ T4	r_k	f_k
l	¬T1 ∧ T2 ∧ ¬T3 ∧ ¬T4	r_l	f_l
m	¬T1 ∧ ¬T2 ∧ T3 ∧ T4	r_m	f_m
n	¬T1 ∧ ¬T2 ∧ T3 ∧ ¬T4	r_n	f_n
o	¬T1 ∧ ¬T2 ∧ ¬T3 ∧ T4	r_o	f_o
p	¬T1 ∧ ¬T2 ∧ ¬T3 ∧ ¬T4	r_p	f_p

We now return to the first problem. We can systematically identify all Boolean expressions by taking combinations of rows from *Table 10.1*, the expression in each row being ORed together. Each such combination will be associated with values of Recall and Fallout, obtained by summing the values of r_i and f_i in the rows, respectively. A value of Precision follows once G is specified. The pair of Recall and Precision values then contributes to the distribution of interest. For example, the boolean search expression:

$$(\mathrm{T1} \wedge \mathrm{T2} \wedge \mathrm{T3} \wedge \mathrm{T4}) \vee (\mathrm{T1} \wedge \mathrm{T2} \wedge \mathrm{T3} \wedge \neg\mathrm{T4}) \vee (\mathrm{T1} \wedge \mathrm{T2} \wedge \neg\mathrm{T3} \wedge \neg\mathrm{T4})$$

will be associated with Recall and Fallout values of $R = r_a + r_b + r_d$ and $F = f_a + f_b + f_d$, and a consequential P value. An algorithm to generate the 'logical surface' of a set of terms is thus:

(1) Read the 2^N pairs of probability values $\{(r_i, f_i) i = 1, 2^N\}$, obtained from an experiment, into a $2^N \times 2$ array. Read a value for G.
(2) Define a threshold value equal to 2^N, with J, an integer variable, initialized to 0.
(3) Define every combination of array rows of size $J+1$. (There are ${}^{2^N}C_{J+1}$ such combinations.) For each combination sum the r_i values to form R and sum the f_i values to form F. Infer P, and place the resulting (R, P) co-ordinate into a cell of a grid defined over $(0,1) \times (0,1)$. This could conveniently be a 60×60 grid, for example.
(4) Increment J by 1, and if $J < 2^N$ return to step 2.
(5) Divide each total of co-ordinates in the cell grid by the grand total of (R, P) co-ordinates, i.e. by 2^{2^N}.

(6) Plot the resulting surface, find the centre of mass, etc.

The algorithm can be modified to delete (R,P) points that arise from combinations involving the all-negated elementary conjunct, as mentioned previously, in which case the grand total changes. We note also that it is doubtful if the algorithm could easily be implemented for N in excess of 4, due to the combinatorial explosion in number of distinct search expressions. (For $N=5$ this number is $2^{32} \doteq 4.3 \times 10^9$.) So what the algorithm produces is the probability distribution on the R–P outcome space when a user of a database selects a search expression arbitrarily, for some given set of search terms.

We turn now to the second problem. Suppose that we restrict attention in the first place to the set of DWFs that each map the 16 elementary conjuncts to 16 distinct real numbers. These DWFs can be distinguished and classified by the rankings (permutations) of the elementary conjuncts that they effect, as previously mentioned, there being 16! classes for DWFs of this type. Our interest is in the hypothetical situation of an enquirer (1) choosing a DWF randomly from one of these classes, and (2) choosing a threshold value randomly from the values 1,2,3, . . . , 16. An algorithm to generate the probability distribution over the Recall–Precision outcome space for this situation (for this species of DWF, for a given instance of information need, and for a given query *qua* set of terms) is as follows:

(1) Define a permutation of the elementary conjuncts. Set J equal to 1. Read G.
(2) Define $R=\sum_{i=J}^{16} r_i$, $F=\sum_{i=J}^{16} f_i$. Infer P.
(3) Put the resulting (R,P) co-ordinate into a cell of a grid defined over $(0,1)\times(0,1)$.
(4) Increment J by 1. If $J<2^N$ then go to step 2 else define a new permutation and go to step 2. (If no new permutation is possible go to the next step.)
(5) Divide each total of co-ordinates in the cell grid by the grand total of (R,P) points, i.e. by $2^N(2^N!)$.
(6) Plot the resulting surface, find the centre of mass, etc.

The surface could be termed the 'document weighting surface' of a set of terms. As noted before, permutations that throw the all-negated elementary conjunct to rank positions other than 1 might be discounted. Also, we could ignore (R,P) points arising from threshold values of 1 if we wished since, unless $r_p=0$, this relates to retrieval of the entire database. (It is recognized as unlikely that an enquirer will require a Recall value of 1.) More complicated algorithms might be identified that produce the Recall–Precision probability surface for DWFs that only weakly order the elementary conjuncts.

The three examples we have given are intended to represent widely different approaches to simulation in information retrieval study. The first was concerned with the 'random number' simulation *technique* serving in that case to show how the effects of choosing different policy options affecting document delivery speed in one type of information environment could be predicted and compared. The second example served to emphasize the necessity of imposing clear definitions on everyday words—in that case 'browsing'—for simulation to be undertaken at all, and perhaps also suggested limitations on the validity of a simulation study when very severely limiting definitions are used. The third example showed how purely formal constructions can usefully be discussed and compared in a particular context using familiar information retrieval concepts, with no additional definition and dealing only in observables.

10.3 Some previous work in simulation applied to information retrieval

For reasons given in the introduction it is in principle impossible to delimit the literature on simulation applied to information retrieval in a satisfactory way. The modelling or representational element, so essential to simulation, is often present in general discussions on retrieval that do not specifically refer to simulation by name. Writers will frequently have used the 'language' of simulation without necessarily having used simulation *techniques* in the narrower sense for exploring the relationships that they discuss, or without having been particularly concerned with optimization of, or intervention in, the process described. We shall adopt as our rather arbitrary criterion for inclusion that systems be formally represented and that relationships between system components be systematically explored using plausible or experimentally-obtained values for the variables involved, with the methodological emphasis on the manipulation of such data. The works meeting this criterion appear to be few in number[20–30]. Gurk's paper[20] is more an indicative description of a prototype of an information retrieval system than a description of a simulation of it. Useful comment directed at simulation work in the general information retrieval context has been offered by Chapman[1] and Salton[12], the latter's monograph reviewing the main models.

The paper by Bourne and Ford[22] is concerned with the *economics* of information retrieval systems. The objective was to estimate the operating cost, and the amounts of equipment and personnel, needed over a given time-period by several hardware information retrieval configurations. Their paper makes the point that on the basis of known data, and a knowledge of the gross characteristics of a proposed system, the costs that *would* be borne in the future can be arrived at by solely manipulative means much more cheaply than by actually building and testing the system, thus underlining one of the basic reasons for undertaking simulation studies. The 'known data' is grouped by them under three headings: 'Time and Cost Data' (wage rates, costs of materials, equipment purchase and maintenance costs, stationery, etc.), 'Statements of Interrelationships' (e.g. item input rate per person, search time per request), and 'Constants' (e.g. amortization period of purchased equipment, interest rate on borrowed capital). Bourne and Ford comment appropriately that the credibility of their type of analysis depends upon the accuracy and completeness of both the analysis of the proposed system and the basic time and cost data, but perhaps they do not sufficiently emphasize the vulnerability of such analyses to rapid technological obsolescence. A further useful point brought out by them is that the *sensitivity* of operating costs (say) or other measures of efficiency or effectiveness, can be explored in a simulation study. (They quote data for annual expenditure as a function of the two independent variables: number of searches per month (from 1 to 100 000), and item input rate per month (from 1 to 100 000 also), to produce an estimated cost range of from $188 000 to $558 000 for one system, and from $166 000 to $551 000 for a second system.)

Baker and Nance[24] report on a study in which the 'system' is defined more generally—so as to include both the users and the funders of the retrieval service—their point being that a more restricted view may lead to suboptimization. (To optimize in respect of system response-time alone, or in respect of unit retrieval cost alone, may be to ignore the costs (or disutility) to the user entailed in (a) noisy (low-Precision) search output, and (b) actual usage of the system, such costs being, possibly, the main causes of low system usage or a poor reputation of it amongst users. (For a related, sceptical viewpoint, see W. S. Cooper[31].) Baker and Nance assume, accordingly, that the funding and operating of a system must be seen as being influenced by user costs and convenience, or utility. The relationships of interest to them are portrayed in two detailed diagrams, and a table of descriptive content. Although the first diagram is a general one, the second, and the table, are relevant to a system having the form of a university departmental library, i.e. to a highly specific system only. The model of the system that is given is moreoever only indicative and no tangible results of the study are given or appear to have been published since.

Reilly's report[26] is unusual in that he was concerned with a *single* user and a *single* type of service, an approach that the Swets model[13,14] can also be interpreted as embodying. Reilly's study assumes that a user estimates both the delivery time of a document from a document-delivery service, and the utility of the service to him prior to making a request from the service. The user's subsequent behaviour is then determined by the truth-values of the inequalities: estimated service time $\leqslant$ need time, and actual service time $\leqslant$ need time, the former being modified with each decision and system response. Since the estimated service time is not an observable in an operational system (though it could be in an experimental environment) the model may not be acceptable to those who insist that simulation should deal only in observables. But non-observables are perhaps acceptable if one can, by assuming their existence and properties, successfully predict observable outcomes using them—the proof of the pudding. Reilly's approach would seem to bring retrieval work closer to the point where user/service interaction is properly heeded and accounted for, as a basis for the fuller system definition needed for the efficient management of information services. The point is in fact made by Reilly (and is also implied in the Baker and Nance paper) that integration of models of different areas of information supply is essential although he does not attempt same. Three such areas or 'levels' are singled out by him in this connection: computer processing centre activities, determination of user behaviour (his main concern), and the delivery of documents. A further point in common between Reilly's report and Baker and Nance's study is that 'a library' should be treated as an information system. Although this has been a commonplace idea in US writings for many years (and is a basis of Salton's recent monograph) there is a still a regrettable reluctance in the UK to view libraries (document supply systems) in the same light as information retrieval services (document *record* supply systems), notwithstanding the common problems each has and the strong interactions that necessarily exist between them. Simulation studies, in offering an abstract 'systems perspective' should help the overcoming of artificial subject barriers such as this.

The work of Blunt, Duquet and Luckie[23] was concerned with the extent of the resources needed, and the response time, of an information retrieval service, and appears to be especially useful in determining the extent to which response time is affected by the competition by queries for resources. Hertz *et al.*[21] and Fried[25] examined the problem of simulating indexed document record files. Heine[29] was concerned with using simulation to predict the effect on Recall–Precision performance of using document age as a component of the query in addition to the more conventional semantic attributes of documents.

Lastly we discuss the two theses by M. D. Cooper[28] and Griffiths[30]. Cooper was essentially concerned, in the simulation part of his study, with the extent to which different queries retrieved different numbers of items from a database. Pseudo-queries and pseudo-documents were defined, each as sets of document attributes. The similarity of a query with a document was expressed as the number of attributes in common between them, i.e. as co-ordination level (to use the Cranfield concept and terminology), and the distribution of the database over non-zero values of the latter was found for a wide variety of queries. Cooper's work appears to be notable for (a) the careful placing of the study in the context of retrieval system evaluation, and (b) the incorporation of term association (i.e. pairwise dependence between terms in some subset of the database, in this case the entire database) in the simulation. However, as stated by him the simulation is limited in its usefulness in that the notion of a partitioning of the database, and in particular a partitioning of the retrieved set, into relevant and non-relevant subsets, is not recognized in the model. The possibility of doing so was rejected by him on the ground that 'not enough information is available to characterize the process' (p. 156). This apparently minor point is dwelt on here because in the writer's view it illustrates the occasional critical dependence of simulation upon *experimental* results (obtained in the laboratory or from operational systems) as well as upon the system description. Griffiths, like Cooper, was concerned with creating pseudo-documents and pseudo-queries in order to simulate the process of post-coordinate searching a database. Unlike Cooper, who chose not to model users' relevance judgements or retrieved sets, Griffiths partitioned the retrieved sets arising in the simulation (identified by a matching process + threshold) by using experimental data obtained from an INSPEC test on retrieval strategies carried out in 1974, and an EEC study of databases containing veterinary literature. The simulation procedure apparently labelled retrieved documents (attached to each co-ordination level) as either relevant or non-relevant on the basis of the value of a Bernoulli variable, but a detailed description of this step and justification of it are unfortunately not given. No attempt was made to model the relevance values of non-retrieved documents (i.e. to partition non-retrieved documents into relevant non-retrieved and non-relevant non-retrieved) so that only Precision, not Recall, is modelled. Co-occurrence frequencies of terms are also not introduced into the model (unlike Cooper's model), presumably because empirical evidence was not available in support of this. Although the main objective of Griffiths was to simulate post-coordinate searching using data obtained from *operational* systems (where Cooper was concerned with more hypothetical data) this objective does not appear to have met with complete success, since (a) operational data for a full validation of the simulation model was not obtainable, and (b) the data that was obtainable from existing small test collections was either inadequate or inapplicable (p. 11).

The main goal of preparing an information retrieval system simulation appropriate to operational retrieval systems appears therefore to be far from complete, since even if a model incorporating valid real data in all significant components could be found, there would still remain the problem of 'designing this in' to a larger model taking into account the motivations of users and supporting agencies, as discussed by Reilly, and Baker and Nance. If valid data cannot be obtained (which seems unlikely), then there is a limitation here *in principle* to the usefulness of the simulation approach.

Further discussion

The general matter of the appropriateness of incorporating test-collection based data into a simulation model deserves careful examination. A prior question is whether research on such collections should be regarded as 'simulation work'—a point touched on in the introduction. Given the immense amount of valuable work that has been done using (say) the Cranfield experimental data[32], by researchers *not* involved in the actual acquisition of that data: for example Salton[3], Sparck Jones[33] or van Rijsbergen and Sparck Jones[34], it seems reasonable to regard such work as 'simulation'. It is oriented towards optimization of a process (information retrieval) through intervention of some kind in it, and is concerned with data already found. It is a *substitute* for further data acquisition achieved through the manipulation of experimental variables, i.e. manipulation of the 'apparatus'. It has become clear from such work that the experimental data incorporated into the test collections are in fact an indispensable input to analyses of a very wide range, as discussed in detail elsewhere in this volume, and that such analyses were not *initially* objectives of the experimental work. They have clearly 'spun off' it. In the sense that such theoretical work uses both the experimental data as input, and uses some of the representations of system components recognized in the experimental work, it is reasonable if somewhat arbitrary to regard it as simulation work. This does not perhaps take us very far, but it points to the conceptual barrier between theoretical work (simulation work in the broader sense) and experimental work as a hazy and rather unsatisfactory one: theory involves constructs; experiment involves apparatus; both manipulate and control data.

Of much greater concern, since it is not just a semantic matter, is the question of whether it is legitimate to use *experimental* data (an archive of which may be stored as a 'test collection') in a simulation study when the data was obtained using questionable methodology. A clear example here is the employment in the experiment of a requirement that 'relevance' be judged in relation to a 'question' where the latter is defined to be a solely verbal construct having no explicit relationship to a context of information need. (The question might be a sentence or paragraph in English, for example.) Here the relevance judgements that appear as experimental data are made with ambiguous terms of reference. Is the arbiter of relevance to *hypothesize* a context of information need before making his decision (which if true would create ambiguity itself—the arbiter's mind may roam over a whole spectrum of possible occasions on which the question could have been put)? Or is the arbiter to look for merely *linguistic* similarity between question and (say) document title? A better experiment would involve relevance judgements being made in a *real* context of information need, with a 'question' appearing not as a term of reference of the judgement but as an articulant serving simply to explore a database in response to an information need. In the latter case it is clearly a *variable,* for a given instance of need. Few phrases have done more damage to the advancement of information retrieval than that of 'relevance to a question' in the author's opinion, in view of the influence on experimental design that that phrase has had, and in view of its inhibiting our awareness of the non-verbal (primitive) basis of relevance decisions. In the writer's view this is simply a consequence of faulty system delimitation. This criticism is destructive of course, and in place of the faulty methodology others more satisfactory than it are in consequence required. *In principle* the correct methodology would be to ask a user to examine all items in a database, and simply note whether each is relevant or non-relevant in relation to a fixed (non-verbal) notion of information need that he says he recognizes. Since this is impracticable for large databases, the two following experimental methodologies might usefully be considered in place of it. The first is to record *behavioural evidence* of relevance decision making, e.g. the records 'used' (in some sense) by the arbiter, or the documents cited in a document written by him. The second methodology is what we term here the 'Virtual Attribute Technique'. This would entail (1) masking *from the search vocabulary* a given term (or other attribute), so that that term becomes virtual or invisible, (2) partitioning the collection to be searched using that term (the relevant set then being the subset of the collection that bears the (virtual) term), and (3) using the remaining search vocabulary in the usual way to try to identify the relevant set so identified. This would have the advantages of objectivity and stability in the relevance-assessments (made *implicitly* by the indexing staff involved in the creation of the database), as well as being consistent with the reality of search-vocabulary development: whereby new terms are introduced with just the purpose of capturing (novel) concepts of relevance. (For relevant stimulating discussion, see Jablonski[35].) There is no reason why this technique could not be applied to real (operational) databases. Techniques such as the Virtual Attribute Technique are urgently needed if simulation work in particular is to develop usefully beyond its present state. They may even serve to diminish our reliance on experimental work as the latter has tended to be construed: as a recording of relevance decision making *et al.* in a laboratory-like environment, rather than as work directed at data gathering in a real 'user/database interaction' environment.

10.4 Conclusions

The work undertaken on the simulation of information retrieval systems appears to have five general features. First it should not be seen in isolation from its natural context—that of information science. It is a mistake to see simulation work solely from a narrow technical viewpoint, since it readily generalizes to a broader systems approach taking in the user's behaviour and even states of knowledge—though such approaches are still in their infancy. Secondly, 'simulation' as a concept, admittedly an ambiguous one, in its broader definitions comes close to being equivalent to the concept or process that we are seeking to describe in information science. Study of that concept has perhaps a unique significance for us. There are intriguing philosophical problems here. Thirdly, simulation work in its narrower technical senses can aid the management of practical operational systems by helping to arrive at good policies on, for example, depth of indexing, optimum online file size, optimum library back-up, and optimum question definition (using 'question' here as a synonym for boolean expression). But the clarifications entailed in describing information systems *usefully* are difficult to arrive at; so difficult that much, if not most, published work on simulation in the area is fairly disappointing. Fourthly, the work undertaken as part of a simulation study in delimiting and describing a system usually improves understanding of the system and suggests hypotheses for further investigation. Lastly it needs to be clearly understood that simulation work, as a species of theoretical work, is always dependent and sometimes critically dependent on valid data from laboratory or operational system studies. Without a continuous stream of such data, reacting symbiotically with simulation/theoretical study, both will be poorer: we will have theory of unproven applicability, and operational systems of unproven optimality. The rationale of the Cranfield experiments must not be lost.

References

1. CHAPMAN, R. L. The case for information system simulation. In: *Proceedings of the Second Congress on the Information System Sciences, 1964*, pp. 477–484, Spartan Books, Washington, D.C. (1965)
2. COOPER, W. S. Indexing documents by gedanken experimentation, *Journal of the American Society for Information Science* **29**, 107–119 (1978)
3. SALTON, G. (Ed.) *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall, Englewood Cliffs, N.J. (1971)
4. BARKER, F. H., VEAL, D. C. and WYATT, B. K. Towards automatic profile construction, *Journal of Documentation* **28**, 44–55 (1972)
5. VERNIMB, C. Automatic query adjustment in document retrieval, *Information Processing and Management* **13**, 339–353 (1977)
6. HEINE, M. H. The 'question' as a fundamental variable in information science. In: *Theory and Application of Information Research*, Proceedings of The Second International Research Forum in Information Science (Ed. O. Harbo and L. Kajberg), pp. 137–145, Mansell, London (1980)
7. VICKERY, B. C. *Information Systems*, Butterworths, London (1973)
8. CHURCHMAN, C. W. *The Systems Approach*, Delacorte Press, New York (1968)
9. MARTIN, F. F. *Computer Modeling and Simulating*, Wiley, New York (1962)
10. GORDON, G. *System Simulation*, 2nd edn, Prentice-Hall, Englewood Cliffs, N.J. (1978)
11. MORSE, P. M. *On Browsing—the Use of Search Theory in the Search for Information*, Technical Report No. 50, Operations Research Center, MIT (1970)
12. SALTON, G. *Dynamic Information and Library Processing*, Prentice-Hall, Englewood Cliffs, N.J. (1975)
13. SWETS, J. A. Information retrieval systems, *Science, New York* **241**, 245–250 (1963)
14. SWETS, J. A. Effectiveness of information retrieval methods, *American Documentation* **20**, 72–89 (1969)
15. ORR, R. H. *et al.* Development of methodologic tools for planning and managing library service: II. Measuring a library's capability for providing documents, *Bulletin of the Medical Library Association* **56**, 241–267 (1968)

16. GORE, D. (Ed.) *Farewell to Alexandria*, Greenwood Press, Westport, Conn. (1976)
17. UNIVERSITY GRANTS COMMITTEE [U.K.] *Capital Provision for University Libraries: Report of a Working Party* (Chairman R. Atkinson). HMSO, London (1976)
18. HEINE, M. H. Measures of language effectiveness and the Swetsian hypotheses, *Journal of Documentation* **31**, 283–287 (1975)
19. ANGIONE, P. V. On the equivalence of boolean and weighted searching based on the convertibility of query forms, *Journal of the American Society for Information Science* **26**, 112–124 (1975)
20. GURK, H. M. and MINKER, J. The design of a simulation for an information processing system, *Journal of the ACM* **8**, 260–270 (1961)
21. HERTZ, D. B. *et al. Research Study of Criteria and Procedures for Evaluating Scientific Information Retrieval Systems*, Arthur Anderson & Co., CFSTI Report AD273115 (1962)
22. BOURNE, C. P. and FORD, D. F. Cost analysis and simulation procedures for the evaluation of large information systems, *American Documentation* **15**, 142–149 (1964)
23. BLUNT, C., DUQUET, R. and LUCKIE, P. Simulation of information systems: some examples, limitations, and an example of a general information systems simulation. In: *Proceedings of the 20th Annual Meeting of the ADI*, Thompson Book Co., Washington, D.C. (1967)
24. BAKER, N. R. and NANCE, R. E. The use of simulation in studying information storage and retrieval systems, *American Documentation* **19**, 363–370 (1968)
25. FRIED, J. B. *et al. Index Simulation Feasibility and Automatic Document Classification*, Technical Report 68–4, Ohio State University, Computer and Information Science Research Center, CFSTI Report PB182597 (1968)
26. REILLY, K. D. *User Determination of Library Request Presentation: a Simulation*, Institute of Library Research, University of California, Los Angeles (1968)
27. REILLY, K. D. *Digital Computer Simulation Models of Library Based Information-retrieval Systems*, Institute of Library Research, University of California, Los Angeles (1969)
28. COOPER, M. D. *Evaluation of Information Retrieval Systems: a Simulation and Cost Approach*, Ph.D. thesis, School of Librarianship, University of California, Berkeley (1971)
29. HEINE, M. H. Incorporation of the age of a document into the retrieval process, *Information Processing and Management* **13**, 35–47 (1977)
30. GRIFFITHS, J.-M. *The Computer Simulation of Information Retrieval Systems*, Ph.D. thesis, School of Library, Archive and Information Studies, University College London (1978)
31. COOPER, W. S. The 'why bother?' theory of information science, *Journal of Informatics* **2**, 2–5 (1978)
32. CLEVERDON, C. W., MILLS, J. and KEEN, E. M. *Factors Determining the Performance of Indexing Systems*, 2 Vols., Cranfield, College of Aeronautics, Aslib Cranfield Research Project (1966)
33. SPARCK JONES, K. Collection properties influencing automatic term classification, *Information Storage and Retrieval* **9**, 499–513 (1973)
34. VAN RIJSBERGEN, C. J. and SPARCK JONES, K. A test for the separation of relevant and non-relevant documents in experimental retrieval collections, *Journal of Documentation* **29**, 251–257 (1973)
35. JABLONSKI, S. The preconstructed vocabulary: a Procrustean bed. *Bulletin of the Medical Library Association* **61**, 21–23 (1973)