(Unpublished sketch written in Aug./Sept 1992; tightened up in Nov. 1993)

'ONLY CONNECT' - SOME THOUGHTS ON DISCOVERY AND CREATIVITY

Alan Macfarlane

PART ONE; ADVICE AND EXPERIENCE OF OTHERS

1. The Copernican/Baconian revolution and its consequences.

The need for the historian or social scientist to think carefully about how he or she arranges material was authoritatively described by the historian and philosopher R.G.Collingwood. In his Autobiography (p.81) he wrote:" I had learnt by first-hand experience that history is not an affair of scissors and paste, but is much more like Bacon's notion of science. This historian has to decide exactly what it is that he wants to know; and if there is no authority to tell him, as in fact (one learns in time) there never is, he has to find a piece of land or something that has got the answer hidden in it, and get the answer out by fair means or foul."

Basically, as Collingwood puts in **The Idea of History**, the historian has become his own authority: "As natural science finds its proper method when the scientist, in Bacon's metaphor, puts Nature to the question, tortures her by experiment in order to wring from her answers to his own questions, so history finds its proper method when the historian puts his authorities in the witness-box, and by cross-questioning extorts from them information which in their original statements they have withheld, either because they did not wish to give it or because they did not possess it." (p.237)

This increasing reliance on the authority of the investigator constitutes a revolution in historical method: "It would be an understatement to say that since 1800 history has passed through a Copernican revolution..." (Autobiography, p.79) He elaborates this revolution elsewhere: "...the Copernican revolution in the theory of historical knowledge has been in principle accomplished. For the common-sense theory, historical truth consists in the historian's belief conforming to the statements of his authorities; Bradley has seen that the historian brings with him to the study of his authorities a criterion of his own by reference to which the authorities themselves are judged. " (Idea, p.240).

The French social historian Marc Bloch wrote roughly at the same time as Collingwood and shared his views. He noted that "It would be sheer fantasy to imagine that for each historical problem there is a unique type of document with a specific sort of use. On the contrary, the deeper the research, the more the light of the evidence must converge from sources of many different kinds". (Craft, 67) This was because "even those texts or archaeological documents which seem the clearest and the most accommodating will speak only when they are properly questioned". (Craft, 69). Elsewhere he commented that "even when most anxious to bear witness, that which the text tells us expressly has ceased to be the primary object of our attention today. Ordinarily , we prick up our ears far more eagerly when we are permitted to overhear what was never intended to be said". (Craft, 63)

This need for cross-questioning also arose, as Collingwood had argued, because, the historian had become a modern scientist. "From the moment when we are no longer resigned to purely and

simply recording the words of our witnesses, from the moment we decide to force them to speak, even against their will, cross-examination becomes more necessary than ever. Indeed it is the prime necessity of well-conducted historical research". (Craft, 64)

All this emphasizes the need for a more thorough approach on the historian's part. The first difficult part is to assemble the appropriate materials. "One of the most difficult tasks of the historian is that of assembling those documents which he considers necessary". (Craft, 69) Equally important is the organization of that material into a form where it can quickly be accessed: "to neglect to organize rationally what comes to us as raw material is in the long run only to deny time - hence, history itself". (Craft, 147)

2. Some case studies.

It is instructive to look at a few of the more conspicuously successful theoreticians and to see how they actually worked. This is likely, if pursued further, to bring some surprises. For instance, as Stibic (Tools,77) writes: "We are surely not surprised to learn that Jules Verne gathered and systematically stored information in the fields of geography, natural science and technology in his well organized collection of 20,000 cards and excerpts. However, it is somewhat unexpected to find, upon visiting the country house of Jack London in Glen Ellen near San Francisco, 188 card-index boxes alongside the writing desk of this writer who had an image of sailor, trapper and adventurer".

For the moment let us look at one representative from three disciplines, history, anthropology and biology. The start of files on some other interesting cases are placed at the end - to be filled up in due course.

Keith Thomas

The current President of the British Academy, Sir Keith Thomas, has written two majestic works; **Religion and the Decline of Magic** and **Man and the Natural World**. In these he gives no clue as to the working methods which enabled him to connect so many facts and interpretations. In an article on 'Ways of Doing Cultural History' (p.78) he briefly describes his method. Firstly he reads widely and deeply: "my only method has been to read as widely as possible in the surviving sources for the period, to soak myself in the society I am studying until I get a feel for it...To achieve this knowledge I try to read everything and look at everything, recognising that some basic assumptions or activities will only be revealed in the most incidental way."

Secondly, while he is reading, he tries to look out for material on a wide range of topics, not his particular present concern. "When reading a particular source, a sermon, say, or a deposition in a church court, I attempt to keep all my different preoccupations in mind and to be particularly alert to incidental revelations unconnected with the document's main purpose. For example, if a preacher, denouncing idleness, compares the futile life of a lazy man to a small child sailing paper boas on a muddy puddle of water, my interest is that here we have evidence that seventeenth century-children played with paper boats."

What he does not explain fully is that he notes all relevant material onto sheets, in microscopic hand-writing. Each observation is given a precise, brief, reference to the source. These sheets are then snipped up with scissors so that the references and quotations become detached from their source order and become a pile of thin slivers of paper. We can now let him take up the account. "I file this away, not on a database, because my habits were formed before the computer came in, but on a piece of paper which I put in an old envelope labelled 'Children's Games'. When the envelope starts to bulge, I tip it out and see what we have got. I then read more systematically so as to fill in the gaps, and after that I may write an article on children's games." Having seen a cupboard half filled with these slivers of papers, and the drawers of old envelopes, I can vouch for the hard work involved.

The final work stage is not described, but I think can be extrapolated from visions of early drafts of **Religion and the Decline of Magic** which Keith Thomas was kind enough to let me see and from my own use of a similar method. When he comes to write an article on 'Children's Games', he would pour out all the hundreds of separate slips and then let his mind play over them so that they began to take shape into sections and paragraphs. He then types out the argument of a paragraph, perhaps a few lines, and attaches the relevant thin pile of apt quotations to it with a paper clip. This is the first draft. In the second draft, the slips are incorporated. We shall find echoes of these methods in other accounts later. It would be interesting to know whether Keith Thomas is running into the unexpected problems of size described in the second half of this piece as his slips mount up and it takes longer to file and retrieve each one by hand.

Claude Levi-Strauss

Let us now consider a part of the work of the great French anthropologist Claude Levi-Strauss. In his several works on mythology he faced a problem of the classic sort. His 'authority' was structured in one way and he wanted to analyse it in another. His material came as a 'text', that is a myth, and he wanted to break it down and re-assemble it in another way. Although there may be an element of self-mockery in the account, this is what he says he did. In an article on 'The Structural Study of Myth'(p.211) he described his method as follows. "How shall we proceed in order to identify and isolate these gross constituent units or mythemes?...The technique which has been applied so far by this writer consists in analysing each myth individually, breaking down its story into the shortest possible sentences, and writing each sentence on an index card bearing a number corresponding to the unfolding of the story. Practically each card will thus show that a certain function is, at a given time, linked to a given subject. Or, to put it otherwise, each gross constituent unit will consist of a **relation.**".

This procedure allows him to connect things which have been separated by the narrative structure of the myth. "Relations, pertaining to the same bundle may appear diachronically at remote intervals, but when we have succeeded in grouping them together we have reorganized..our myth according to a time referent of a new nature, corresponding to the prerequisite of the initial hypothesis..."

He quickly recognizes that the huge number of slips generated by this method soon leads to technical problems which require some kind of computer solution (pp.228-9). "At this point it seems unfortunate that with the limited means at the disposal of French anthropological research no further advance can be made. It should be emphasized that the task of analyzing mythological literature, which is extremely bulky, and of breaking it down into its constituent units, requires team work and technical help. A variant of average length requires several hundred cards to be properly analyzed. To discover a suitable pattern of rows and columns for those cards, special devices are needed, consisting of vertical boards about six feet long and four and a half feet high, where cards can be pigeon-holed and moved at will. In order to build up three-dimensional models enabling one to compare the variants, several such boards are necessary, and this in turn requires a spacious workshop, a commodity particularly unavailable in Western Europe nowadays. Furthermore, as soon as the frame of reference becomes multi-dimensional (which occurs at an early stage, as has been shown above) the board system has to be replaced by perforated cards, which in turn require IBM equipment..."

Presumably what he is referring to is the primitive holerith cards of the 1960's, when he wrote the article. But this is leading in the direction of cross-tabulation and coding, which is what one wants to avoid. The whole approach would be different now with small and powerful computers.

Charles Darwin

It is not surprising, in view of his achievements, that considerable attention should have been paid to Darwin's working methods.

One account is given in **Life and Letters of Darwin** (i,82) as follows. Darwin kept "from thirty to forty large portfolios, in cabinets with labelled shelves, into which I can at once put a detached reference or memorandum. I have bought many books, and at their ends I make an index of all the facts that concern my work; of, if the book is not my own, write out a separate abstract, and of such abstracts I have a large drawer full. Before beginning on any subject I look to all the short indexes and make a general and classified index, and by taking the one or more proper portfolios I have all the information collected during my life ready for use." One really needs a photograph to gain a picture of this - and no doubt one is available (q.v.)

A more detailed account is given in a study of Darwin's marginalia by De Gregario. (see Mario A. Di Gregario et al., **Charles Darwin's Marginalia,** vol. 1, pp.xii-xiii. We are told firstly of the way in which he read. "During the basic reading, intensive or otherwise, the margin is scored and peppered with comments. At the end of the reading, he would now list out the location of his more important comments and margin-scores on an inside cover (usually the back cover), occasionally adding brief mnemonic notes. " This conforms with his own description - though we should add that he also noted onto separate sheets if the book was not his own.

We then move onto the stage when he is going to write. "The book will now probably lie fallow - maybe even for a number of years - until the lucky moment arrives. At this stage, the list of locations is re-examined, and a new, shorter, list made on a separate sheet of paper of the most important locations, now with details in the form of long-hand notes about the information to be gleaned at those locations...We have the feeling that he hardly ever reread the book itself...". This is a gloss on the passage above in which Darwin explains how he makes a condensed index.

We are now at the stage of writing. "The vital slips containing the vital gems at this point reach the prime of their working lives: we imagine those relevant to the publication in progress now collected in a heap (...or pile) on the writing table, being finally reviewed. (some marked for quotation, others not) on the slips or at original locations. The set of slips, together with CD's own notes and drafts, combine for a while into 'Portfolios of working notes' for the writing of the publication in question. Once the publication has been pieced together, "slips all put in proper places" (572h) in other words they are stuck for any future reference usually inside the back cover of the now fully-harvested book. A slip may take part in this 'cycle' a number of times - its important underlying content, as we shall see later, being the broad theoretical themes invoked by the data recorded on it".

Di Gregario then try to penetrate deeper into the way Darwin's mind actually worked: "a summary of our hypothesis about CD's main mode of 'processing' scientific reading matter: the margins, end-notes and the slips of various different paper types constitute physically discrete strata or layers, corresponding more or less closely to different bouts of attention. Insofar as these bouts imply an accumulation across different 'layers' of time, the metaphor of geological deposition seems quite reasonable." As Di Gregario continues, "In fact the 'layers' concept begins to unlock the inner nature of CD's mode of working with sources: and, indeed, we should ideally look upon the whole great corpus of marks and comments not piecemeal, but as a **single** complex laminate - fused layers not only of time and attention, as we have seen, but also of types of **response** to the source material, and also layers of **themes** reflecting CD's lifelong theoretical preoccupations".

Although there are still puzzles, this has taken us one step further into the creative process. To proceed further down this difficult path, perhaps we can take a detour by way of Koestler's work on creativity.

3. Connections and creativity.

There are ample indications that those who have most enriched our view of the world have done this through making curious and unexpected connections. Writing of Weber, for instance, Collins (Weberian, 7) states: "He is full of unexpected insights and subterranean connections, although many of them have proved too forbidding for most readers attempting to dig through the unaccustomed historical examples..." The external manifestations of this process are described by de Tocqueville (Memoir,2,339): "When I have gathered in this toilsome harvest, I retire, as it were, into myself; I examine with extreme care, collate and connect the notions which I have acquired, and I make it a rule to give the result, without bestowing a thought on the inferences which others may draw from what I write".

How does the mind think it is working in these moments of inspiration? A few classic accounts are worth citing. One is by the geneticist Galton. "There seems to be a presence-chamber in my mind where full consciousness holds court, and where two or three sides are at the same time in audience, and an ante-chamber full of more or less allied ideas, which is situated just beyond the full ken of consciousness. Out of this ante-chamber the ideas most nearly allied to those in the presence-chamber appear to be summoned in a mechanically logical way, and to have their turn of audience". (quoted in Koestler, Creation, p.160) This is one of the most powerful metaphors for that groping on the edges of consciousness which seems always to be present in great science and great art.

Another hint of the process, with a similar idea of a core and margins, is given by the psychologist William James. "The great field for new discoveries is always the unclassified residuum. Round about the accredited and orderly facts of every science there ever flows a sort of dust-cloud of exceptional observations, of occurrences minute and irregular and seldom met with, which it always proves more easy to ignore than to attend to". (quoted in Koestler, Creation, p.191)

The need to connect is further described by the great French mathematician Poincare in a famous analysis. "Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart...Most combinations so formed would be entirely sterile; but certain among them, very rare, are the most fruitful of all". (in Koestler, Creation, p.164) This passage looks similar to, but is not identical to, another valuable piece by Poincare in Nadel (Foundations, 243), which is also worth quoting: "The experimental method is intended to 'reveal unsuspected relations between....facts, long since known, but wrongly believed to be unrelated to each other. Among the combinations we choose, the most fruitful are often those which are formed of elements borrowed from widely separated domains." His thoughts would be worth following up further.

The combinations occur from shaking and juxtaposing. "'It is obvious', says Hadamard, 'that invention or discovery, be it in mathematics or anywhere else, takes place by combining ideas...the Latin verb 'cogito' for 'to think' etymologically means 'to shake together'. St. Augustine had already noticed that and also observed that 'intelligo' means 'to select among'". (quoted in Koestler, Creation, 12)

Koestler's own observations on the creative process are worth noting. Creativity "...is signalled by the spontaneous flash of insight which shows a familiar situation or event in a new light, and elicits a new response to it. The bisociative act connects previously unconnected matrices of experience; it makes us 'understand what it is to be awake, to be living on several planes at once' (to quote T.S. Eliot, somewhat out of context)." (p.45) This process "does not create something out of nothing; it uncovers, selects, re-shuffles, combines, synthesizes already existing facts, ideas, faculties, skills". (120) There is a "sudden interlocking of two previously unrelated skills, or matrices of thought". (p.121) Or, to put it another way, "All decisive advances in the history of scientific thought can be described in terms of mental cross-fertilization between different disciplines". (230) We begin to see similarities where before there was only difference. "The most important feature of original experimental thinking is the discovery of overlap and agreement where formerly only isolation and

difference were recognized".(p.232)

How does one achieve this? One element is the necessity to forget. It is important to be able to forget what we already think we know. "To undo wrong connections, faulty integrations, is half the game. To acquire a new habit is easy, because one main function of the nervous system is to act as a habit-forming machine; to break out of the habit is an almost heroic feat of mind or character. The prerequisite of originality is the art of forgetting, at the proper moment, what we know". (p.190)

Another element is the placing of our material in another context:"...the art of handling the same bundle of data as before, but placing them in a new system of relations with one another by giving them a different framework, all of which virtually means putting on a different kind of thinking-cap for the moment". (235) One way of achieving this is to force ourselves into new challenges. "But to recapture the erstwhile magic in all its freshness, he must turn to something new; experimental theatre, avant-garde films, or Japanese Kabuki, perhaps; novel experiences which compel him to strain his imagination, in order to make sense of the seemingly absurd - to participate, and re-create". (p.336)

There may be even more practical advice. Darwin had his 'thinking path' along which he would stride, trying to sort out his ideas. In this he was demonstrating Max Weber's view that: "Both, enthusiasm and work, and above all both of them jointly, can entice the idea. Ideas occur to us when they please, not when it pleases us. The best ideas do indeed occur to one's mind in the way in which Ihering describes it: when smoking a cigar on the sofa; or as Helmholtz states of himself with scientific exactitude: when taking a walk on a slowly ascending street; or in a similar way. In any case, ideas come when we do not expect them, and not when we are brooding and searching at our desks. Yet ideas would certainly not come to mind had we not brooded at our desks and searched for answers with passionate devotion." (From Max Weber, 136).

It may even be that some dulling of the conscious brain is needed, as the poet Houseman described (see whole section in Koestler, 318)."Having drunk a pint of beer at luncheon - beer is a sedative to the brain, and my afternoons are the least intellectual portion of my life - I would go out for a walk of two or three hours". The writer then becomes 'possessed', Einstein's 'rape of the mind', or, as Kipling described it "My Daemon was with me in the Jungle Books, Kim, and both Puck books, and good care I took to walk delicately, lest he should withdraw. I know that he did not, because when those books were finished they said so themselves with, almost, the water-hammer click of a tap turned off....Note here. When your Daemon is in charge, do not try to think consciously. Drift, wait, and obey." (Autobiography, 210)

4. Longer accounts of the method.

As Weber realized, we need a combination of a passionate search for answers, and then moments of relaxation and calm. How can one help entice the Daemon and make his life less strenuous when he does arrive. We here begin to turn to some of the practicalities of the research method, taking up the thread of 'one fact one card' which we began to explore in the second section.

One classic account of the method which we have seen was alluded to by all of our various authors, some kind of indexing which liberated material from its original context and allowed the mind to re-arrange it, is described by Beatrice Webb.

a. The Webbs

In the well-known appendix to **My Apprenticeship**, Beatrice Webb wrote: "It is difficult to persuade the accomplished graduate of Oxford or Cambridge that an indispensable instrument in the technique of sociological enquiry - seeing that without it any of the methods of acquiring facts can seldom be used effectively - is the making of notes". (pp.426-7)

The method of writing 'one fact on one card' which she described "enables the scientific worker to break up his subject-matter, so as to isolate and examine at his leisure its various component parts, and to recombine them in new and experimental groupings in order to discover which sequences of events have a causal significance". (427)

The liberating effects of this shuffling of paper are well described. "To put it paradoxically, by exercising your reason on the separate facts displayed, in an appropriate way, on hundreds, perhaps thousands, of separate pieces of paper, you may discover which of a series of hypotheses best explains the processes underlying the rise, growth, change or decay of a given social institution, or the character of the actions and reactions of different elements of a given social environment". (427)

Physical details are important. For instance, a standardized size of card, and placing the same information in the same place on the card, it makes it easier to move quickly through the materials. "Thus, a carefully planned 'display', and, above all, identity of arrangement, greatly facilitates the shuffling and reshuffling of the sheets, according as it is desired to bring the facts under review in an arrangement according to place, time or any other grouping". (340) For instance, "By adopting our method of one sheet for one subject, one place and one date, all the sheets could be rapidly reshuffled in chronological order; and the whole of our material might have been surveyed and summarised exclusively from the standpoint of chronology". (432)

The result of this mechanical device was that it was possible to look at questions in numerous different ways: "By the method of note-taking that I have described, it was practicable to sort out all our thousands of separate pieces of paper according to any, or successively according to all, or these categories or combination of categories..." (432). It also, most importantly, provoked clashes and surprises. "Not once, but frequently has the general impression with regard to the causal sequence of events, with which we had started our enquiry, or which had arisen spontaneously during the examination of documents, the taking of evidence or the observation of the working of an organisation, been seriously modified, or completely reversed, when we have been simultaneously confronted by all the separate notes relating to the point at issue". (433)

Beatrice Webb again stresses the surprising effects on creativity of this apparently simple strategy. "I realise how difficult it is to convince students - especially those with a 'literary' rather than a 'scientific' training - that it is by just this use of such a mechanical device as the shuffling of sheets of notes, and just at this stage, that the process of investigation often fertile in actual discoveries". (433) This is partly because "Most students seem to assume that it is the previous stage of making observations and taking notes which is that of discovery." (433) If one used the metaphor of film, the editing stage, which is when one assembles the material into a new order, is widely known to be as important as the filming or 'collecting' stage. Yet most people do not realize this when undertaking literary work.

Wright Mills

Again in an appendix, this time to his work on **The Sociological Imagination** C.Wright Mills gives a number of further lucid descriptions of how he works. He describes, in words somewhat akin to Darwin's, how "After making my crude outline I examined my entire file, not only those parts of it that obviously bore on my topic, but also those which seemed to have no relevance whatsoever. Imagination is often successfully invited by putting together hitherto isolated items, by finding unsuspected connexions. I made new units in the file for this particular range of problems, which, of course, led to new arrangements of other parts of the file." (221)

What happens is that "As you rearrange a filing system, you often find that you are, as it were, loosening your imagination. Apparently this occurs by means of your attempt to combine various ideas and notes on different topics. It is a sort of logic of combination, and 'chance' sometimes plays

a curiously large part in it. In a relaxed way, you try to engage your intellectual resources, as exemplified in the file, with the new themes." (221)

Into the original files go all sorts of things, made possible by a very flexible storage system which "... encourages you to capture 'fringe thoughts': various ideas which may be by-products of everyday life, snatches of conversation overheard on the street, or, for that matter dreams. Once noted, these may lead to more systematic thinking, as well as lend intellectual relevance to more directed experience". (217) These files contain "ideas, personal notes, excerpts from books, bibliographical items, and outlines of projects". (219) This constitutes an ever enriched resource. "Then as you pursue your work you will notice that no one project ever dominates it, or sets the master categories in which it is arranged. In fact, the use of the file encourages expansion of the categories which you use in your thinking". (p.219)

Even the actual method of note-taking forces you to think about what you are doing. "Merely to name an item of experience often invites you to explain it; the mere taking of a note from a book is often a prod to reflection". (219) The file is "a continually growing store of facts and ideas, from the most vague to the most finished". (220) When one comes to write, it is really a development from these files: "the idea and the plan came out of my files, for all projects with me begin and end with them, and books are simply organized released from the continuous work that goes into them". (221)

The files lead to those unexpected associations and connections which we have seen are the essence of true discovery. In intellectual work, "...there is an unexpected quality about it, perhaps because its essence is the combination of ideas that no one expected were combinable - say a mess of ideas from German philosophy and British economics". (233) The unexpectedness comes from the method of proceeding - the mind has broken apart and is now able to re-combine elements. "On the most concrete level, the rearranging of the file, as I have already said, is one way to invite imagination. You simply dump out heretofore disconnected folders, mixing up their contents, and then re-sort them. You try to do it in a more or less relaxed way". (233) The classifications created by the original materials can be broken. "Many of the general notions you come upon, as you think about them, will be cast into types. A new classification is the usual beginning of fruitful developments". (234) Thus the researcher, "Rather than rest content with existing classifications, in particular, common-sense ones, you will search for their common denominators and for differentiating factors within and between them. Good types require that the criteria of classification be explicit and systematic. To make them so you must develop the habit of cross-classification". (234)

Mills realizes that linking things, connecting, or as he calls it 'cross-classification' is the key to discovery. "For a working sociologist, cross-classification is what diagramming a sentence is for a diligent grammarian. In many ways, cross-classification is the very grammar of the sociological imagination". (235)

What he has described has much overlap with the other descriptions. You extract, abstract, cross-relate, and re-integrate into new patterns. A final description gives an abstract description of what Darwin described. "After you decide on some 'release', you will try to use your entire file, your browsing in libraries, your conversation, your selections of people - all for this topic or theme. You are trying to build a little world containing all the key elements which enter into the work at hand, to put each in its place in a systematic way, continually to readjust this framework around developments in each part of it". (245)

We now have a few hints of what needs to be done. But how, in practice, is one to do it? Here one turns from the general accounts to autobiography. In what follows I will describe what I have tried to do, first through the development of various systems of hand-indexing and then, in a preliminary way, in relation to computerized indexing.

PART TWO; SOME INDEXING STRATEGIES

1. Hand-indexing.

The systems of hand indexing which Sarah Harrison and I have developed over the year are numerous. One of them is a very large index of about 45,000 small cards on numerous topics arranged into a hierarchical index system. The principle of hierarchical indexing is the only practicable one for such large sets of data. Its general form is described by Stibic. "Classification is based on a preconceived plan: the whole field of interest is a priori divided into a number of classes. Nearly all classification systems are hierarchical, i.e. the whole field is divided into a few main classes, each of these is subdivided into subclasses, and this process of splitting-up can be repeated down to the required detail". (p.97) In fact the hierarchical divisions were not worked out all at once, but gradually evolved, the main headings in the first five years, and then lower level ones over the next fifteen years.

In this 'topics' database, as I now call it, the top level of the hierarchy has about twenty major headings, each with up to 2000 cards, under titles such as:

AGRICULTURE, MEDICINE, ECONOMY, MARRIAGE, POLITICS, **CAPITALISM**
and so on. Each one is roughly a drawer of small cards.

Within each heading would be the next level; for instance, under CAPITALISM would be 'CAUSES OF CAPITALISM', 'CONSEQUENCES OF CAPITALISM', **FEUDALISM AND CAPITALISM**, ETC.
These are divided in my system by a green card.

At the third level of the index, would be further sub-divisions, for instance under the above would headings such as : 'WAS ENGLAND FEUDAL?' 'IS EUROPEAN FEUDALISM UNIQUE', 'WHEN DID FEUDALISM END' and so on. Each of these would be divided by a pink card.

At a rough guess, therefore, there would be some two thousand cards per unit of the top level, each of these might be divided into ten, with two hundred cards, and each of these divided into ten, with an average of twenty cards. In fact, there is enormous variance.

At the level of the card, there would be a local heading, for instance on the top of the card there might be:

'Feudalism and Contract'
 There then would be the abstracted text or quotation or idea, written in the middle of the card, for instance:

"The master who taught us that 'the movement of the progressive
societies has hitherto been a movement from Status to Contract' was quick to add that feudal society was governed by the law of contract. There is no paradox here..." (referring to Henry Maine)

There would then be the reference to the source of the idea or quotation, for instance,

author: Maitland
title: English Law, 1
page: 233

The size of the cards, half a standard 5" by 3" record card, ie. two and a half inches by three, forces one to be brief. The quotation above, with under 40 words, is about the average; almost always the

text would be between ten and fifty words. The maximum would be about 100 words. This is one respect in which the method somewhat differs from that advocated by the Webbs, who clearly envisaged longer extracts.

This hand-indexing method was adapted for use in the study of particular communities, historical and anthropological, by changing the hierarchy. In essence the top level became:

PERSON, PLACE, DATE, SOURCE

and the material was then indexed in fairly conventional ways under these. (for a longer description, see **Reconstructing**).

The method was used to write my D.Phil. thesis and book on Witchcraft and later works on Marriage, Individualism, Population etc. It performed all the functions which previously quoted writers allude to. It brought together surprising information, it stimulated comparisons, it enabled one to be adding to a dossier on a broad front etc. For about fifteen years, as it grew from a few cards in 1963, as I started my D.Phil, to the time when I completed **Individualism**, when it included about 35,000 cards, it worked very well. It then ran into problems, of which for a long time I was only dimly aware, but are now, hopefully resolved.

As long as one keeps to about 20,000 cards and does not want to change the hierarchical classification, the hand-indexing and hierarchical system is probably manageable. After about 1980, however, I began to be aware that, as with everything else, the law of diminishing marginal returns was starting to set in. This was due to various features which I have not found widely discussed in the arts and social sciences.

One of these concerns the basic fact of size. There would appear to be a general law of indexing that the larger the body of other cards into which a card has to be put, the slower it becomes to file a specific card. Thus, if one has one drawer with 2000 cards, one can probably put a card straight into it. By 1980, after about 17 years of indexing, I had about 17 drawers with about 35,000 cards. It had thus become necessary when filing the new index cards to sort things first into major headings, then sub-headings, and only then to approach a drawer to look at the specific headings. It might be taking up to 30 seconds to 1 minute to put away a single card - and it was very tedious and quite strenuous work. The difficulty did not just arise from the fact that there were more and more places where it could be filed, because there were more cards. It also arose from the nature of hierarchical classification systems.

As one put away index cards, one was trying to remember the layer upon layer of previous classifications one had established. Did one have a category for attitudes towards menstruation, or should one make a new one? Should one put it under 'Gender', 'Life Cycle', 'Sex', 'Body' or what? The difficulties can be imagined.

When I refer to the layers of classification, this refers to a related difficulty, namely that over a period of twenty years one's understandings and hence classifications shifted. As one's interests shifted, old distinctions and divisions meant less, and new one's were created. One effect of this, was that to file something under a system whose principles were laid down 20 years ago, one had to try to think back into the classifications of that time - another cause of friction and difficulty.

The overall classification itself became less meaningful, yet, being hierarchical, while it was possible to tinker with it, it was impossible to change it fundamentally. As Stibic points out, "The principal disadvantage of a hierarchical classification system is its rigidity. It is difficult to adapt the existing divisions, and it is practically impossible to change its basic structure". (98)

This problem was related to another, which is rather briefly glossed over in Mills' account of

'cross-classification'. By definition, all 'cards' contain a **relation** between at least two things, and probably many more,. Hence each card needs to go under **at least** two headings. In fact, it should probably go into three or four places in the filing system. If we take the example of the quotation from Maitland quoted above, that could have been indexed under:

Maitland's ideas
Maine's ideas
feudalism - nature of
contract - relation to feudalism
status - destruction of by feudalism
progress - evolutionary views of
modernity - what caused it

and no doubt other headings as well. It has been arbitrarily assigned to one of these (feudalism) and tends to be lost to all the others.

This necessity to choose only one, or occasionally two, places where material could be filed, adds to the problems of finding the information again when one came to work on a topic. In other words, the law of diminishing marginal returns also applied to information retrieval. This is partly just the problem of more cards to search through. There is a tension because, in theory, the larger the data base (i.e. the more cards), the more exciting and interesting the searches should prove. This was often true and perhaps reached a peak when the number of cards had reached about 30,000.

Yet it also began to be obvious that it often takes a very long time to find relevant information. As the hay-stack got larger, it naturally became more difficult to find the few relevant needles. What was relatively simple with one thousand became progressively more difficult.

This simple problem of numbers was exacerbated by the often implicitly and little recognized shifting of classifications over time. A card with the same words written on it might have been put in all sorts of places depending on the time in which it was filed away. In order to find it again, or to know under which subjects to look for information on a particular topic, one often had to remember when about one might have indexed something, the classification system at that time and then make the search.

Yet all of this ran counter to another aim, namely the desire to encourage overlap, change classifications, cross-classification. Without pursuing this goal, how was one going to break out of the mould of previous classifications?

The result of all these un-analysed pressures was that after about assembling about 40,000 or so cards, I began to lose heart. It was usually quicker to remember an author and go to the book - to rely on human memory. One might then use the method which is described by Darwin, namely of indexing a book at the back, and as long as one could remember the author, one could find the material. This is the method which Christopher Hill the historian described to me as using. One indexes the book in terms of one's interests at the front or back, and just goes to it. If one does not have the book, one abstracts it and does the same.

It is probable that this is about as far as one can go using the unaided human memory and bits of paper. Perhaps some people could master more than 40,000 cards - but they would still find that putting away cards and retrieving them became less and less easy over time. It is at about this time, in the 1980's, that the development of desk-top database systems emerged as a possible solution to the problem.

6. The computer solution.

Although no doubt the same ultimate laws apply to computers, for a while at least the use of computers, like any new technology, gives the user the feeling that he or she has temporarily suspended the laws of diminishing marginal returns.

During the last few years, the data gathering methods in the arts and social sciences have developed rapidly. In particular the development of photography, film, tape-recording and other gathering devices, as well as, recently, the use of new input devices (scanning) and laptop computers have increased the amount of material that can be gathered. Yet the even more important activity of analysing and absorbing and re-classifying the materials, which should absorb two thirds of the effort and time, has received less attention.

Let us break this down into the various stages. The first is the filing away of what we had previously envisaged as written record cards, and now become separate computer 'records'. With the development of modern computing, this problem is largely solved. This is despite the fact that the mathematical rule which explains why it becomes impossible for a hand card index to expand much beyond about fifty thousand cards still operates. This rule, when applied to computers, states that the larger the set of records existing in a database, the longer it takes the computer to index/store a new item. The mathematics of this rule is well laid out on p.275 of PC Magazine, July 1992, where it is shown why "a database of any kind will gradually slow down as more data is added", because "to place a record in an index, an efficient program can do no better on average than a time proportional to the natural logarithm of the number of items already in the index." This means, for instance, that if there is only one index item in the database, it will take 0.001 seconds to insert one item. If there was no slowing down, one thousand items should be inserted in 1 second. In fact, because of the above law, it takes 7 seconds. And so it goes on.

The reason why the threshold is raised, of course, is because of the speed of computers. Not only are they already very fast, but their central processing power is growing so fast, that it is easily keeping pace, indeed exceeds, the needs of most social scientists. Secondly, the strain is on the machine. Though it may take half an hour in 'real time' to add a batch of records to a database, the human being only takes a few seconds to set up the process and can then go and do more interesting things. Thus the problem of adding record is, at least temporarily, solved.

The same relaxation of the rules is true of the other operation - finding the data again. While it is true that the larger the data set, the longer, in principle, it takes, to find any particular item, this difficulty is overcome by two features of computers. Firstly, their actual processing power is again increasing so fast, that, as with adding material, it is far exceeding the social scientist's or historians requirements. Secondly, with the growing sophistication of database management systems and information retrieval software, which have moved from hierarchical to relational to probabilistic retrieval in a little over a decade, with increasingly sophisticated synonym systems, suffix stripping, thesauri, semantic analysis and other developments in information retrieval, it is actually getting easier and easier to find things.

Both of the problems, filing away material and then finding it again are linked to the problem of hierarchical classifications. Such classifications were necessary when indexing by hand because some variant of hierarchical classification is always necessary in hand systems - whether in a museum, library, or any other classification of data. That is how human beings find things - divide them, and then sub-divide again and so on in some form of hierarchical taxonomy. But, as we saw, this is frustrating in several ways. Firstly it tends to be inflexible over time; secondly it forces arbitrary decisions as to where to put items of information; thirdly it has the very effect of preventing interesting connections, because one is just imposing yet another pattern. Such a pattern is probably a common-sense one, based on one's training, and reflecting the very conventional classifications which one is trying to overcome. Although it will bring things together which are held apart in other books, in the end it still seems to hold things apart as much as linking them.

In some ways, the hand method suffers from the problem of most retrieval systems, namely that one has to know what questions one will ask when one puts in the information. Yet, by definition, one cannot anticipate, and should not be able to, what questions one will want to ask as further interactions with the world occur.

It would appear that one example of a new database system based on probabilistic statistics, for instance the CDSi system which we are developing, helps overcome this problem. This system just takes raw strings of characters and allows one to put in any shape or pattern one likes to find them.

In the Maitland quotation used above, one could still have found the material by familiar methods. One could have put in the query by name, title, or major heading 'feudalism and contract' and found the material. But more interestingly, one could be working, say, on the problem of the peculiarity of western capitalism and put in the query 'to what extent did feudalism represent a movement from status to contract'. If one searches on the words 'status contract feudal' in a database of 47,000 the system found the above quotation in less than two seconds (as the second best answer to the query). What is more, one can browse through a number of other interesting records on the same theme, for instance a discussion by Granet of the same phenomenon happening in China.
The real delight is that, since one is now pursuing problems and associations, one travels through all the artificial divisions created in the hand system and finds material of an unexpected kind in unexpected places.

One had always sensed that half-forgotten treasures were in one's material, gathered over a long period but like a squirrel's nuts buried in autumn, half forgotten. One knew, for instance, if one was trying to consider the use of animal analogies in politics that there might be material under all sorts of fields - under general headings of 'animals', 'metaphors and similes', 'politics' etc. But also in other places - say under a section headed 'nature and culture' or 'Hobbes' or 'Greek thought' or whatever. Only a small part of what was relevant would be found under the specific heading under which, twenty years ago, one would have decided to put it. Now, thanks to a brainless, senseless, statistical, machine, all sorts of weird and wonderful connections can be made to refresh the mind and stimulate the imagination.

As well as the advantages of speed, the computer has the not-inconsiderable one of compactness. In order to undertake really deep research, one needs one's material compactly around one. As Stibic advises "The basic rule for designing a functional work place is: Keep your work space as small as possible! The ideal is that all tools, papers, documents, communication equipment and frequently used devices are within arm's reach". (30) Yet, here again, the law of diminishing marginal returns sets in - as the paper in the form of books, notes and index cards increases, it tends to move away from one and occupy larger and larger spaces.

Even if one partially overcomes the problem in a particular work place, much research is done 'on the move', on holidays, in evenings, in all sorts of situations. Once one has 40,000 cards, one is anchored to a reasonable sized study, unless one just takes out one section - abandoning all the other relevant material. To a certain extent, information is enriched just by the process of concentrating it. The fact that one can carry in one's brief case the 40000, or, in future, ten times that number, of cards and find any one of them, is a considerable benefit.

Yet all this does assume a good retrieval system. The problem is well demonstrated by straight boolean retrieval. If one uses the conventional database and put in a query of the and/or/not query for 'status or contract or feudal', in my database one gets some 1128 answers. The particularly interesting ones, like the one above, would probably, on average, come about half way through - i.e. one would have to look through several hundred answers before finding the particular quote one was looking for. On the other hand, if one tried to compel a more interesting answer by asking 'status and contract and feudal', the answer would have been found more quickly, but all sorts of other interesting materials would not emerge. Each time one narrows the query down with a boolean

query, one loses interesting material, but to leave it 'open' makes searching very tedious.

The system of 'probabilistic' searches, combined with stemming or stripping off the suffix, seems to deal with this effectively, since it will give the 'best' sort of answer first, and then present other answers in decreasing order of probability of interest. This makes it exciting to look through, bringing to the front what one is looking for. A system which is purely statistical gives the **feeling** of being intuitive.

There are, of course, traps and dangers as there are with any new technology. There is a danger that one becomes obsessed with the method and think that it will solve problems. It is therefore necessary to remember the usual warnings. These can be listed very briefly.
i. Garbage In Garbage Out.... If one puts in uninteresting material, one will get out uninteresting facts.
ii. Dangers of a sort of elevated 'scissors and paste' of the kind Collingwood warned against - a 'clipboard' to which one affixes all one's favourite quotes.
iii. One will find false correlations and spurious connections - because there is so much in the system (equivalent to statistical fallacy of cross-tabulating everything against everything).
iv. The 'problems' are not generated by the system, but by the human being and must come from outside. Hence one needs to think of interesting questions - generated by other methods (e.g. the comparative method, q.v). This system does, however, help to refine and suggest new lines of investigation, by connecting things and suggesting new connections through the method of "query expansion".
v. One of the advantages of cards was that they could be easily shuffled into any order, thrown in the air and assemble into new patterns etc. This is not yet available in this system, where after a search the 'hits' tend to fall into a sequential file. Of course, one could format them into little card shapes, or slips and print them out, cut them out, and then shuffle them about. This is something to think about, as it should be possible to do on the screen, as well as manually.

APPENDIX; OTHER POSSIBLE CASE STUDIES

Over time I hope to collect material on the working methods of others who have left revealing accounts of how they creatively connected materials. These might include

literary figures:
Coleridge: (see the excellent accounts in Lowes, 'The Road to Xanadu' which is full of insights) also Charles Dickens
For instance, Loren Eisely writes in **Darwin and Mysterious Mr X** (p.82), "The tendency has been to accept 'the sacred river', the sunless seas of dream as the primary source of Coleridge's inspiration. Out of these misty depths, according to entranced critics, were drawn in poetic ecstasy fragments of traveller's les transmuted forever in the subconscious mind of the poet". As Lamb, quoted on p.89 puts it, "the true poet dreams being awake".

Proust: I believe, though I have not checked, that he kept a very complex card index system in order to write 'A la recherche...' - and so it would be worth looking into, and also his remarks on memory and association. For instance, he wrote interestingly, concerning the relational method. "One can place indefinitely in succession, in a description, the objects which figured in the place described: truth will not begin (to appear) until the moment when the writer will take two different objects, will place them in a relationship, analogue in the world of art to the unique relationship of the causal law in the world of science". (quoted in Lienhardt , Anthropology , p.153.

anthropologists:

Sir J.G.Frazer: some of Frazer's notebooks are in the archive of the Department of Social Anthropology in Cambridge. They show very clearly how he worked with a 'one fact one card' indexing system, using perforated slips and carbon paper so that one copy would be kept in source order, and the other could be indexed under headings.

local historians:
Possibly an extract from the work of Alan Macfarlane and Sarah Harrison on Re-constructing historical communities. In our work on reconstructing English historical records, we developed a hand method based on the idea of bringing together people, places, etc. by means of a 'one fact one card' method. It may be worth briefly alluding to the descriptions of this in 'Reconstructing Historical Communities'.

   The basic aim of the approach adopted here is to gather together and analyse all the records which relate to a certain set of individuals in the past....(p.37)..the essence of the approach is the necessity that several different records bear on a particular individual at different points in his or her life. This concept will be familiar to those who have studied the 'family reconstitution' technique..." We alluded to Bloch's remarks about the difficulty yet necessity of assembling relevant material and wrote: (p.81) "We must include under 'assemble' not only the gathering of records into one place, but the further indexing and analysing of these records in such a way that they may more easily be used. Both these processes may appear to the non-historian to be fairly trivial matters, yet upon the methods of collecting and preparing the data for subsequent analysis will depend the whole success or failure of a project concerned with studying a particular community. Again quoting Bloch on the need to organize the gathered material rational, we continued "the material must be broken down in various ways by means of indexing so that it becomes practicable to ask complex questions. The records as they exist in the archives were not organized to help the social investigator."

   This was elaborated a little later (p.83). "Efficient indexing is the essential research tool for any really fruitful study of the kind envisaged here. The success or failure of historical reconstruction will largely depend on the sophistication and thoroughness of the indexing. When attempting to solve problems, it is absolutely essential that one should be able to move very quickly along a series of links, a chain of names or events or places, in order to see whether some hypothesis is correct. If such movement is very arduous or time-consuming, the tenuous thread will be lost...the final aim of an investigator of past communities attempting to study similar problems is to build his information around him in such a form that he can approach as closely as possible to the social scientist's privileged position..."