T.R. NO. (1)

Technical Report

No. 1

ORGANIZATION OF KNOWLEDGE
AND INFORMATION RETRIEVAL

by

A. Ghose & A.S. Dhawle

~~1973/1~~

WP 1973/1

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

ORGANIZATION OF KNOWLEDGE AND INFORMATION RETRIEVAL

by

A. Ghose & A.S. Dhawle

ORGANIZATION OF KNOWLEDGE AND INFORMATION RETRIEVAL.

(Ghose and Dhawle)

Because of the advent of computers and their almost unlimited capabilities of information processing, we want to design a theory and a system of arrangement of information relevant for management so that information retrieval for the users is optimal. Most of such existing systems classify information into mutually exclusive classes corresponding to the traditional organization or classification of knowledge, like arts, science, engineering, medicine. The interdisciplinary development of knowledge has made the traditional organization of knowledge into mutually exclusive subjects or disciplines obsolete. Subjects like biophysics and bioengineering show how easily the boundaries of subjects are crossed! Sometimes some isolate ideas become full fledged subjects, like management, development studies, policy sciences. Ranganathan's facet concept of a subject, as a train of fundamental characteristics such that any other characteristic is expressible as a joining or linking of the fundamental characteristics, remained restricted to bibliographic classification, i.e., arrangement of documents in the library. The focus of our research is to design a system of non-bibliographic classification of knowledge directed towards the information retrieval needs of the users. Researches in this type of non-bibliographic classification are being carried out in different parts of the world to lead towards a new philosophy of organization of knowledge.

Development of a computer system corresponding to this theory for the information stored at Vikram Sarabhai Library would be a long term project, and involve continual updating. In order to develop a computer system, keeping in view the constraints of computer limitations, financial resources, and the time, we would like to restrict our project to one functional area of management, i.e., marketing.

However our theory would be so general that the general system of notation would be independent of feasible computer systems, and can be operated manually.

A classification of the universe of knowledge, technically known as the organization of knowledge, was developed even in the ancient times. Aristotle's basic classification influenced organization of knowledge during the middle ages. Although Leibniz is also indebted to it in some way, he modified it, seeing how the faculties of universities of his time were organised. The universe of knowledge was divided into mutually exclusive classes, because it was felt that certain forms of knowledge had nothing in common with other forms of knowledge. Thus Bacon's division of knowledge into History, Literature, Philosophy, was based on the understanding that the nature of knowledge in each of the classes is different from one another. Each of the classes correspond to a particular faculty of human mind i.e. memory, reason, imagination. Thus according to Bacon, "History" is knowledge of facts corresponding to the faculty of the mind - memory. Literature is knowledge of creation of man's imagination. Philosophy guided by reason encompassed a wide range of subjects including Law, Political Science, Mathematics, Logic, etc. Natural Science was included in Philosophy because of its rational approach in the method of observations, experiementations, and generalizations by inductive logic. That Bacon's classification was too broad, and certain subjects within such broad classes, should be independent classes, was felt by one of his French contemporaries Naude' who gave the classification[1] : (i) Theology, (ii) Medicine, (iii) Jurisprudence, (iv) History, (v) Philosophy, (vi) Mathematics, and (vii) Polite (Belle) Literature. In this scheme Mathematics, Jurisprudence and Literature are recognised as independent disciplines.

Library classification systems which are in use today like, Dewey Decimal Classification (DC), Universal Decimal Classification (UDC), Library of Congress Classification (LC), Bibliographic Classification (BC) are based on the theories of organization of knowledge developed during the 19th century. These theories accepted the traditional idea of dividing the universe of knowledge into mutually exclusive classes. Because of rapid development of science and technology, the problem was to determine the classes and their sub-classes (to determine the generic and specific) of a particular field of knowledge satisfying the specialists. The subjects within a class should be homogeneous. These theories tried to solve such problems.

The arrangement of documents on the shelves in a library is according to main subjects, where documents of subsubjects are put together. A corresponding notation system for the documents is developed which determine the location of the documents. This sort of classification of the information stored in a library is technically known as bibliographic classification (3, p. 365), because the primary objective of such a classification is to distinguish the location of documents within the library. The following extract from UDC classification will make this clear.[5]

6	Applied science
62	Engineering
624	Civil engineering
624.1	Substructure
624.13	Earthwork
624.131	Soil mechanics
624.131.4	Soil properties
624.131.43	Physical properties
624.131.439	Tensile properties
624.131.439.4	Compressive strength

If one is searching for information on "Compressive strength of soils", then first of all it is necessary to know to which main class it belongs. Once the main class is identified as "Civil Engineering" one has to follow through the subsclasses of "Sub-structures", "Earthwork", "Soil mechanics", ...etc. to find "Compressive strength of soils". The decimal notation reflects the structure of subclasses within subclasses. Thus 624.131.439 represents a class which is included in the class represented by 624.131.43 which again is included in 624.131.4 and so on.

The above system of classification assumes however that each particular topic can be considered as a subtopic of a more general topic which again can be considered as a subtopic of a more general topic and so on until one reaches the main subject. The problem of classification however is not so simple. The philosophical and linguistic problems which arise in connection with classification of knowledge have been clearly shown in the article of Richmond (3;pp.364-365) see Appendix A. Sometime it is difficult to decide what is the main topic and what is a subtopic of it. For example one would normally consider "the concept of limit" as a subtopic of the main topic "Calculus". On the other hand if one considers the abstract concept of limit in terms of topological spaces "Calculus" can be considered as a subtopic of "the concept of limit".

Another problem arises when a concept is related significantly to more than one concept, though to none of them in the relationship of genus and species, i.e. general and specific. Consider for example the relationship of "advertisement" to "profit", "attitude", "product", "ethics". The inadequancy of "genus-species" relationship as a basis for library classification, commensurate with the modern

development of knowledge, was felt even in the nineteenth century, However the first systematic attempt to develop a new system of library classification was made by Ranganathan in 1925 who introduced the concept of facet analysis. Whereas in the earlier enumerative systems the principal of classification of a subject is based on Aristotle's notion of genus-differentia-species, facet analysis is based on breaking up a subject into certain number of terms which can be considered as the "categories" of the subject, and then building up the subject by linking and joining these terms, like one assembles the different parts of a mechano set to construct a structure. Consider for example the topic "treatment of lungs tubercolosis by X-ray" in Medicine. The information relevant to this topic cannot be found out from a library which has an enumerative system of classification. "Tubercolosis" would occur as a subtopic of "Problem (Diseases)", but then "treatment by X-ray" would not be considered as a subtopic of "Tubercolosis". By facet analysis if one considers "Problem (Diseases)", "Handling (Treatment)", "Organ (Parts of the Human Body)" as three categories of Medicine, then the information relevant to "treatment of lungs tubercolosis by X-ray" can be represented as "Medicine (Lungs) : (Tubercolosis) : (X-ray treatment). Just as the UDC represents a system of notation to represent classification based on the idea of going from general to specific, similarly Ranganathan's system of notation, represents the joining or interlinking of different facets of a subject.

In implementing the facets concept for library classification one faces two major problems. (i) One has to determine the facets of a subject in such way that it is acceptable to all users. Ranganathan's breaking up of a subject into the facets does not take into account users' conception about the subject. (ii) the system of notation has to be simple enough so that users can interpret the artifici

language easily and guide themselves to find the relevant information. Ranganathan's Colon Classification is much too complex to expect the users to learn it and then make use of it for literature search. Because of these two disadvantages the system of Colon Classification remains to be of use only to the librarian as a means of bibliographic classification, and not to the user seeking a particular information. If information is considered as the set of all documents that can exist at a given moment of time, then the system of Colon Classification gives an efficient classification of information, for even new subjects that may crop up from time to time can be embraced within this notation. However Colon Classification system cannot be regarded as system of organization of knowledge, if knowledge is information as interpreted by the users.[4]

The aim of our research is to develop a theory of organization of knowledge relevent for management studies keeping in view the different perspectives of the information retrieval needs of the users. In determining the facets of the subject "Management" we are guided by the different perspectives in which users approach the subject. Ranganathan's facet concept was restricted to categories within a subject. He did not take into account the "fusion" of two or more subjects leading to a new discipline. Management is a discipline where many subjects are used. It is essentially interdisciplinary, as are also many new subjects like "Policy Sciences", "Planning", "Development Studies", "Manpower Planning". This is illustrated by the attached Appendices B - F. One of the objectives of the research is to extend the facet concept to an interdisciplinary subject like Management and then develop a system of notation so that users can easily understand it, and make use of it for information retrieval. This will mean a significant improvement compared to the present system of subject-cataloguing which gives the user a very vague idea of where he might find

the relevant information. The next step would be to develop a computer system, taking into account all the hardware limitations, which would enable the users to get quite specific information almost in no time.

In designing the facet system for Marketing we first of all determine from a survey of literature on Marketing, the different subjects which are used. These are some of our categories like Quantitative Methods, Economics, Politics, etc..... (The present list is not fully comprehensive as shown in the Appendix G). Within each of the categories different topics are included. For the present demonstration we have selected only a few topics. For instance, under the category 'Psychology' we have Motivation, Image, Learning, etc. Besides the main subjects as categories we have two other categories (i) general concepts (ii) specific concepts of Marketing. Specific concepts of Marketing are those which are peculiar to Marketing as a discipline, like "Market Segmentation", "Market Research", "Distribution Channels", etc. General concepts are those which can be applied to any specific concept of Marketing or even to any other concept of the subjects given as categories. For example "Strategy", "Ethics", "Policy" are general concepts. The user can combine these concepts in any way he wants. Thus he may combine "Policy", "Planning", "Development", "Product". In the present system however the user would get the same class of literature if he had given "Development, Policy, Planning, Product". This shows that the language used for information retrieval is at the moment very elementary; just words, separated by commas. A more complex language, where prepositions and logical connectives are used, would enable the user to express the topic of his interest more precisely, and thus the corresponding computer system would be able to serve him much more adequately and efficiently. Even the present system

of notation used for computer programming is not optimal. Much work needs to be done in developing a system of notation similar to but simpler than Ranganathan's system which can be operated manually. The development of an optimal computer system corresponding to this would be a problem by itself. The manual system and computer system should complement each other. "It will be a generalized information retrieval language, with vocabulary and grammar, suitable for general and specialised information centers and for manual and mechanical retrieval systems."[2]

The nature of this research itself is interdisciplinary. Beginning from Library classification it goes into theories of organization of knowledge, logic, semantics, theory of formal languages, and ends with a system of notation and a non-numerical system of computer programming.

REFERENCES

(1) Bliss, Harry Evelyn. The organization of knowledge in Libraries and the subject-approach to books. 2nd ed. rev. New York, Wilson, 1939. p.194

(2) Foskett, Douglas J. "Classification Research Group, 1952-1968". In Encyclopedia of Library and Information Science, Vol. 5. New York, Mercil Dikker, 1971. p. 145

(3) Richmond, Phyllis A. "Reading List in Classification Theory", Library Resources and Technical Services, Vol.16, No.3, Summer 1972.

(4) Shera, Jesse H. The foundations of Education for Librarianship. New York, Wiley, 1972, p. 117

(5) Vickery, B.C. Techniques of Information Retrieval. London, Butterworths, 1970. p.107

APPENDICES

A Extract of Richmond's article : Richmond, Phyllis A. "Reading List in Classification Theory", Library Resources and Technical Services, Vol.16 No.3, Summer 1972, pp.364-366

B Management as an Integrated Discipline - Figure : Urwick, L.F. "Why the so-called "Classicists" Endure", Management International Review, Vol.11, 1971/1, p.13

C A Systems Overview of Marketing Management - Figure : Lazer, William and Kelley, Eugene J. Interdisciplinary Contributions to Marketing Management. East Lansing, Michigan, Bureau of Business and Economic Research, College of Business and Public Service, Michigan State University, 1959, pp.20-21

D Measurement in Marketing - Figure : Lazer, William and Kelley, Eugene J. Interdisciplinary Contributions to Marketing Management. East Lansing, Michigan, Bureau of Business and Economic Research, College of Business and Public Service, Michigan State University, 1959, p.18

E A Systems View of Sales Forecasting - Figure : Lazer, William and Kelley, Eugene J. Interdisciplinary Contributions to Marketing Management. East Lansing, Michigan, Bureau of Business and Economic Research, College of Business and Public Service, Michigan State University, 1959, p.24

F Manpower Studies and Related Disciplines - Figure : Institute of Manpower Studies and London Graduate School of Business Studies, London. A classification for Manpower Studies. 1972, p.2

G A System of Categories for Marketing - Table : Prepared by the authors of this paper.

Appendix A

OUTLINE

I. Definition; what is classification?
- A. Classification as description
- B. Classification as division
- C. Classification as hypothesis
- D. Classification as scientific method
- E. Classification as a convenient ordering device
- F. Classification as coordination or grouping
- G. Classification as folk taxonomy
- H. Classification as an umbrella
- I. Classification as a structure
- J. Classification as systematic concept coordination
- K. Classification as a common sense arrangement

II. Bibliographic classification
- A. General
- B. Kinds of systems
  - i. Universal classification systems (summary)
  - ii. Enumerative systems
    - a) Dewey Decimal Classification
    - b) Universal Decimal Classification
    - c) Library of Congress Classification
    - d) Bibliographical Classification
  - iii. Faceted classifications
    - a) Colon (summary)
    - b) Classification Research Group types
      - (1) General
      - (2) Specific
      - (3) Later
      - (4) Bibliography
    - c) Faceted other than Classification Research Group types

iv. Other

III. Types of nonbibliographic classification

IV. Recent views of classification

A. Criticism and goals

B. Purpose and method

C. Eclectisism: some ideas adopted by various classificationists

i. Integrative levels

ii. General systems theory

iii. Other

D. Toward a philosophy of classification.

V. Subjects related to classification

A. Notation

B. Linguistic approaches: terminology, language, linguistics, content analysis, stylistics

C. The subject approach, including latent structure in indexing

i. Subject headings, descriptors, index terms

ii. Thesauri

D. The classified catalog

E. Automatic indexing and classification

F. Automation of classification procedures and results

i. Input (as part of cateloguing automation)

ii. Input/output (as a system of subject analysis)

VI. Miscellaneous background material

A. Scientific method, philosophy

B. Information science, nee documentation

C. Cataloguing codes

D. Logic

E. Statistical methods

F. System analysis

G. Communication theory, symbolism

H. Evaluation

I. Computers and computer programming

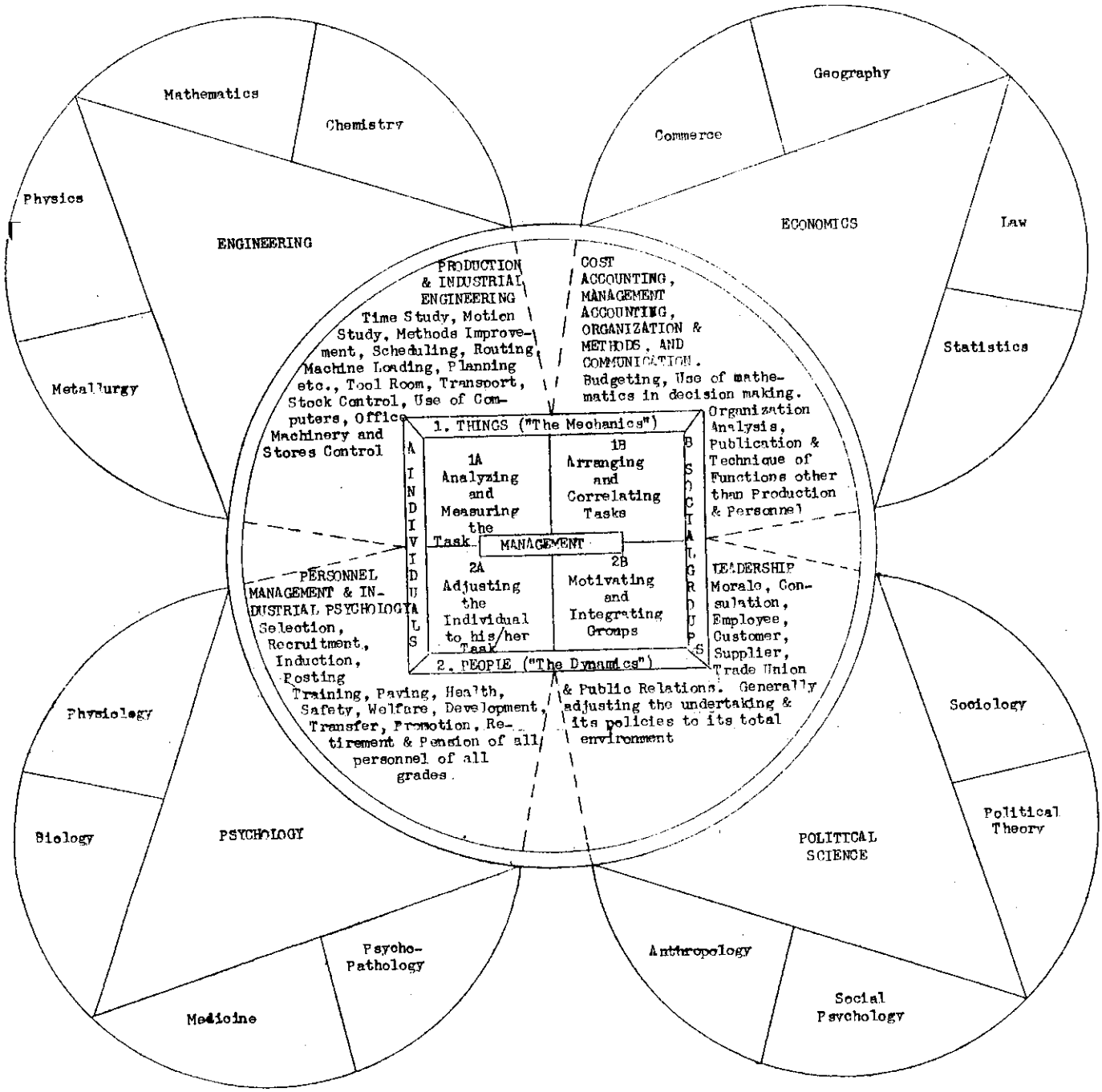
J. Surveys, the future

K. Library automation

Source: Richmond, Phyllis A. "Reading List in Classification Theory", Library Resources & Technical Services, Vol. 16 No.3, Summer 1972, pp. 364-366.

Figure I

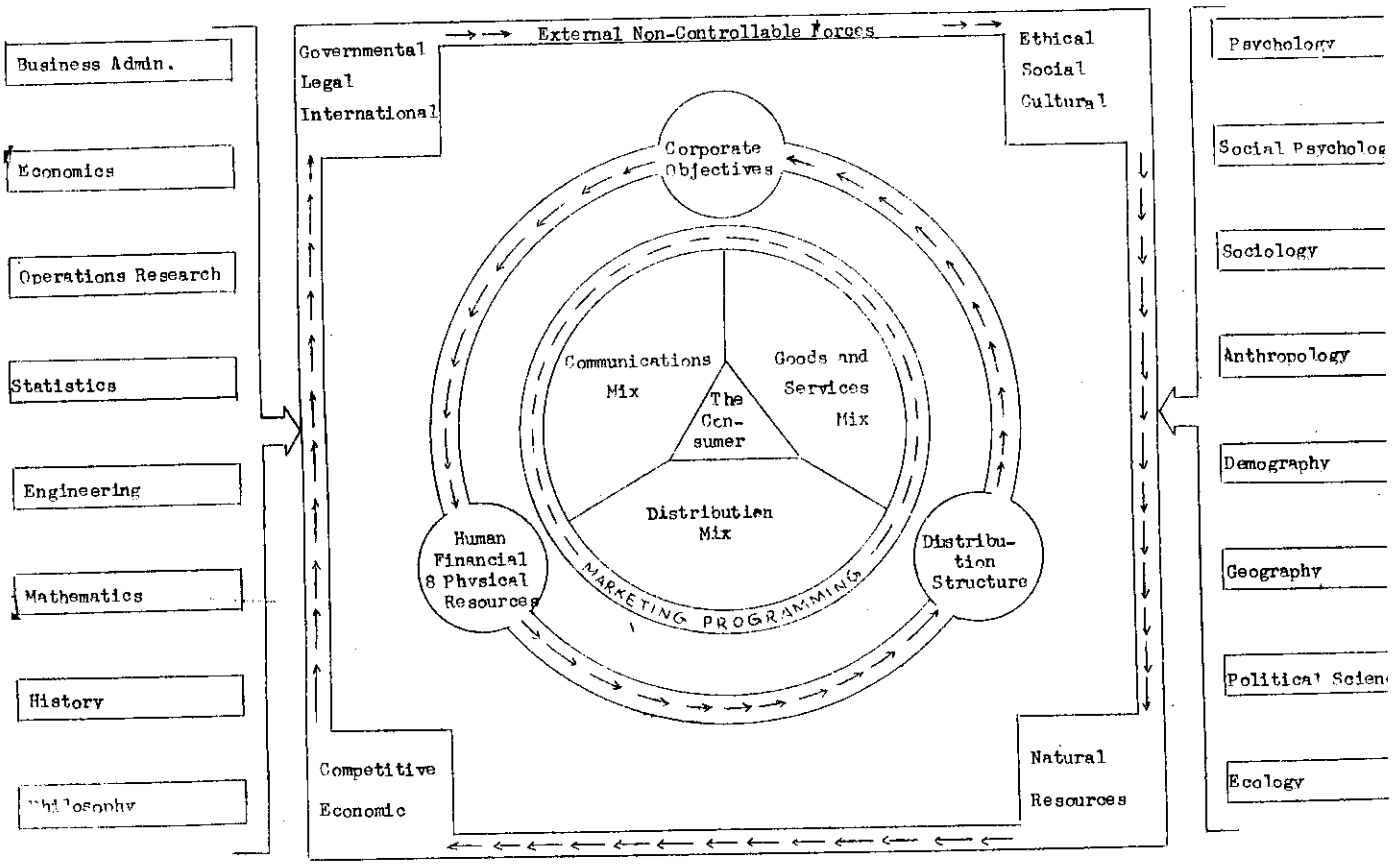
Management as an Integrating Discipline



Source: Urwick, L.F., Why the So-called "Classicists" Endure, Management International Review, Vol. II, 1971/1, p. 13.

Figure II

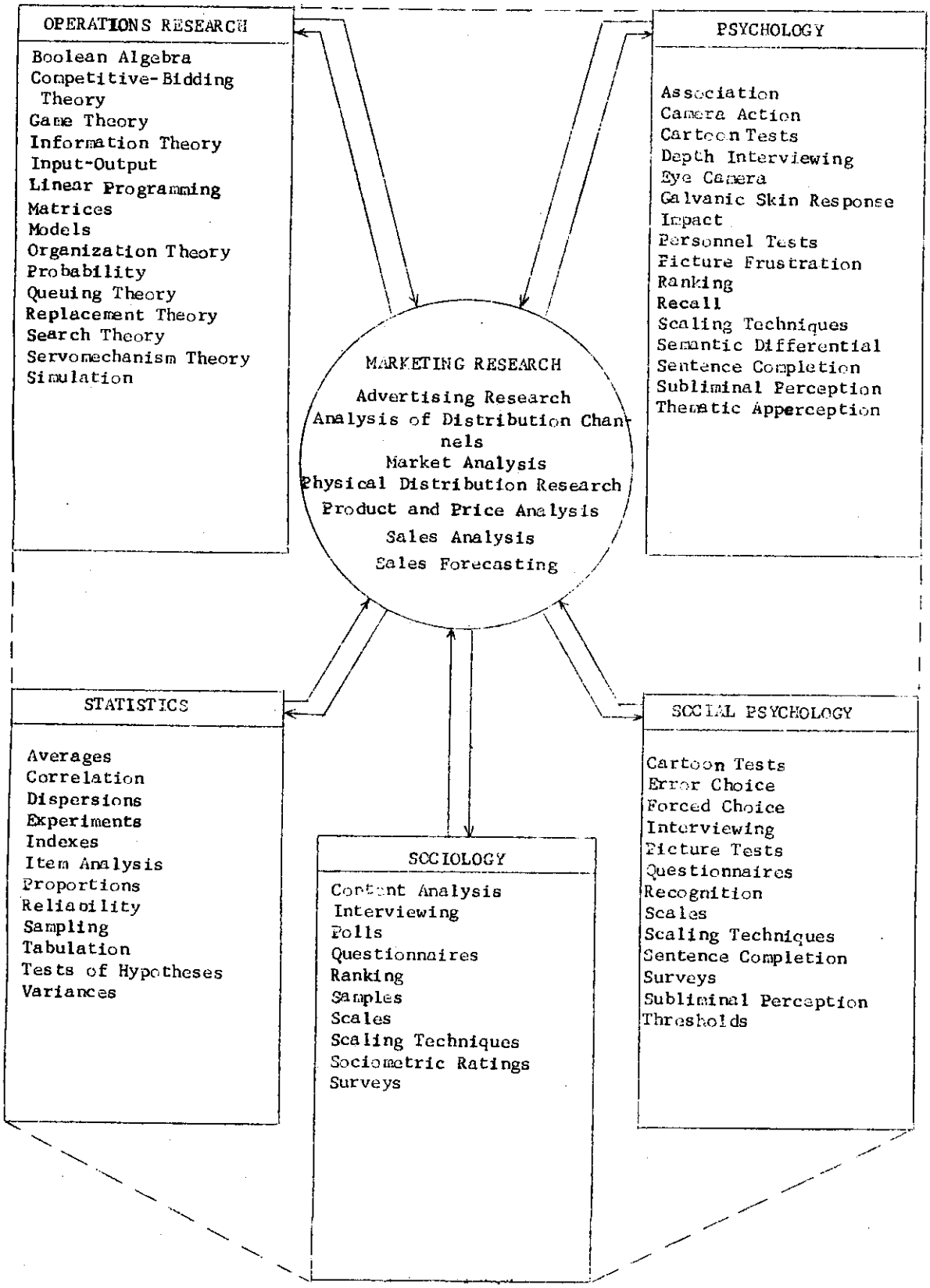
A Systems Overview of Marketing Management



Source: William Lazer and Eugene J. Kelley. Interdisciplinary Contributions to Marketing Management, Michigan State University, Michigan.

Figure III

Measurement in Marketing

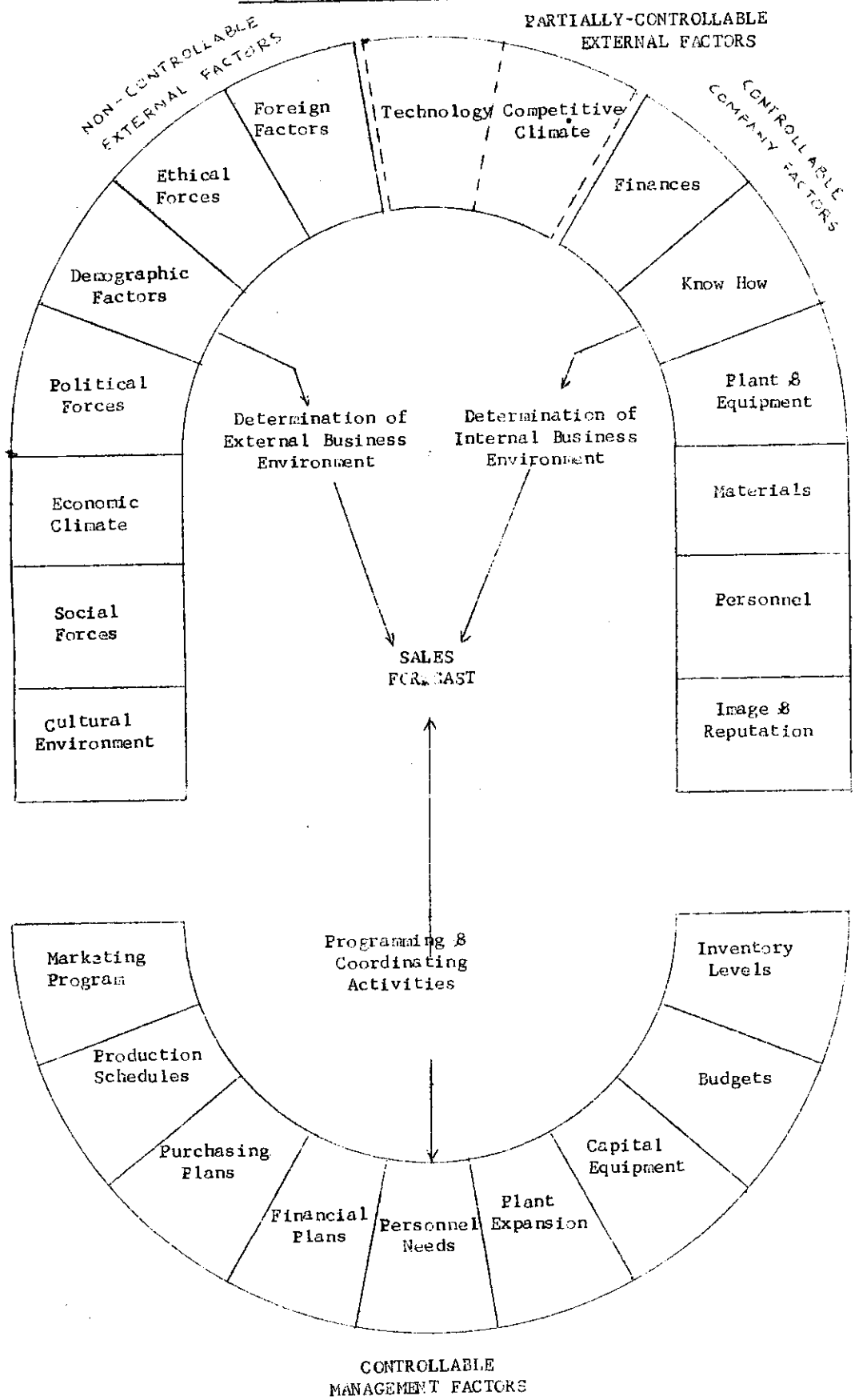


Source: William Lazer and Eugene J. Kelley. Interdisciplinary Contributions to Marketing Management, Michigan, Michigan State University.

Appendix E

Figure IV

A Systems View of Sales Forecasting

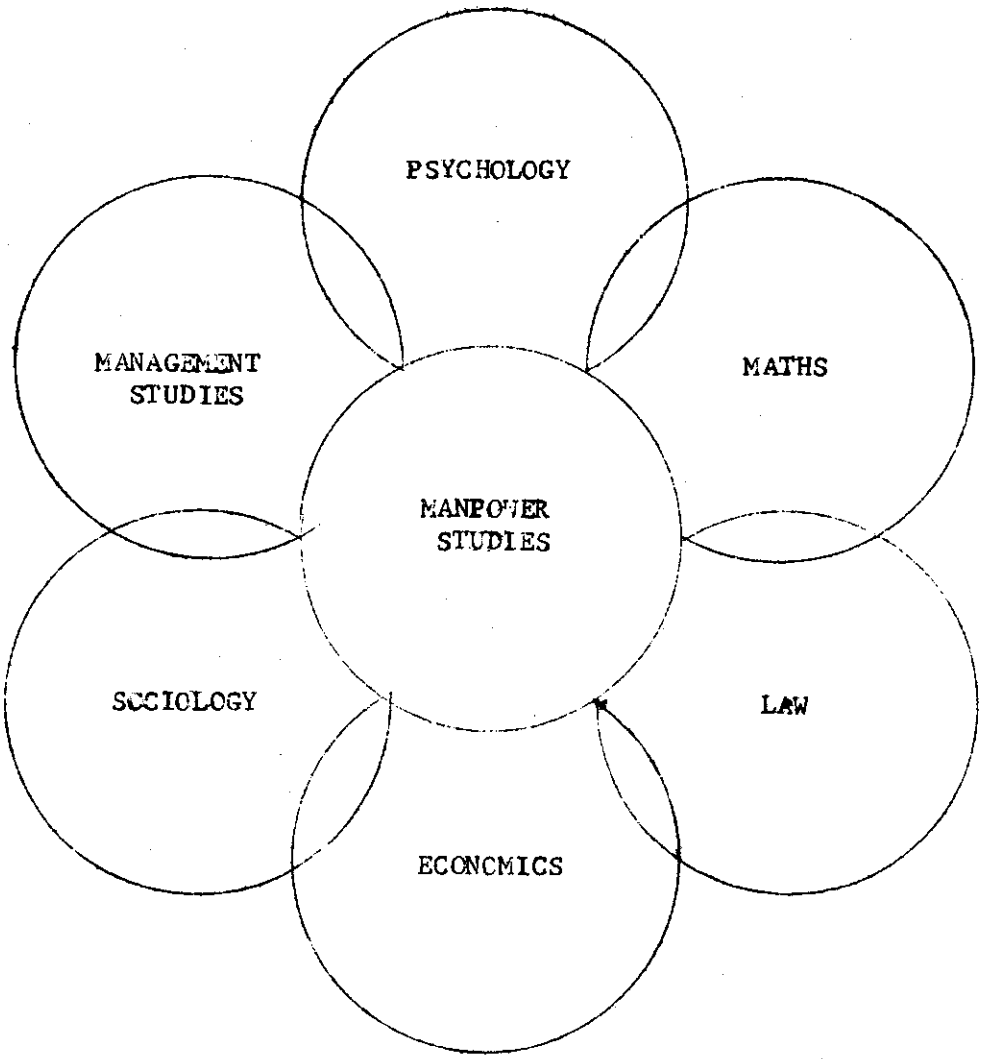


Source: William Lazer and Eugene J. Kelley. Interdisciplinary Contributions to Marketing Management. Michigan, Michigan State University.

Appendix F

Figure V

Manpower Studies and Related Disciplines



Source: A Classification for Manpower Studies, Institute of Manpower Studies and London Graduate School of Business Studies, London, 1972, p. 2.

TABLE : A SYSTEM OF CATEGORIES FOR MARKETING*

Appendix G

Quantitative Method/QM	Finance & Accounting/FA	Economic/EC	Anthropology/AP	Sociology/SI	Psychology/PS	Political Science/PL	General	Marketing
Linear Programming	Break Even	Cost-Benefit	Culture	Family	Attitude	Government	Agricultural	Agencies
Optimisation techniques	Cost	Forecasting		Group Behaviour	Image	Political Groups	Control	Brand
Simulation	Budget	Price		Social Classification	Learning	Power	Decision	Consumer
Statistical Methods	Credit	Profit			Loyalty		Ethics	Consumer Behaviour
					Motivation		Fashion	Distribution
					Perception		Forecasting	Manufacturer
							Industrial	Market Organisation
							Information Systems	Market Segmentation
							Innovation	Market Mix
							Law	Physical
							Life Cycle	Price
							Management	Product Mix
							New	Products
							Objective	Retailing
							Obsolescence	Sales
							Planning	Services
							Problem Solving	Wholesaling
							Policy	Packaging
							Strategy	
							Systems approach	
							Technology	
							Testing	
							Total Systems	
							Values	

* The group of Categories is chosen here keeping in view the users' convenience.

Source: Prepared by the authors of this paper.

No. ....1.....

To

Chairman (Research)
IIMA

Im lewis

Technical Report

Title of the report	:	ORGANISATION OF KNOWLEDGE AND INFORMATION RETRIEVAL
Name of the Author	:	Prof. A Ghose & Anand S. Dhawle
Under which area do you like to be classified?	:	Production and Quantitative Methods Area (Management and Information Systems)

ABSTRACT (within 250 words)

This paper examins the traditional organisation or classification of knowledge based on mutually exclusive subjects or disciplines which have become obsolete. It attempts to give a new look and to develop a classification system for knowledge rather than for knowledge embodied in documents. Therefore, the focus of research of this paper is to design a system of non-bibliographic classification of knowledge directed towards the information retrieval needs of the users or utilisation of knowledge. It also analyses the theory (fundamental concepts) gone into classification of the universe of knowledge technically known as the organisation of knowledge in ancient times, and also, Library classification schemes designed to classify documents embodied with subjects and disciplines on traditional lines. It probes into the fundamental concepts of organisation of knowledge, and utilisation of hardware (Computers and their unlimited capabilities for information retrieval). To start with, this research work has tried to develop a theory of

organisation of knowledge relevant for management studies (specially Marketing, a pilot test done to show the feasibility by using Institute's Computer), keeping in view the different perspectives of the information retrieval needs of the users to demonstrate the inter-disciplinary approach. It identifies the areas of further research related to the subject of this paper e.g. Theories of organisation of knowledge, Logic, Sementics, Theory of formal languages, System of notation and a non-numerical system of computer programming, etc.

Please indicate restrictions if any that the author wishes to place upon this note. .....

.....

Date: July 20, 1973

Signature of the Author
(Anand S Dhawle)

ASD/nksk