RM-78-47

"PROBLEMS OF SCALE" - THE CASE FOR IIASA RESEARCH

M.F. Cantley
V.N. Glagolev

September 1978

Preface

Why "Problems of Scale"?

There are several reasons why this topic is a suitable and important one for study at IIASA. In the first place, problems of scale are real issues about which a large number of people in almost every country of the world are concerned. They are thus "universal" problems. Certain aspects of "global" problems may also be viewed in terms of scale effects. Articles about real problems of scale appear in the literature of many industries and activities - electricity generation, hospitals, coal mines, super tankers, chemical plants, steel plants, aluminium, regional planning, governmental decentralization, etc.

Such problems appear to have common features, but are usually tackled separately, without very much borrowing from previous studies. Some of these common features are as follows:

(i) a traditional economic model, embodying relationships between size, performance and cost, with the resulting historic trend being towards increasing size for maximum economic advantage;

(ii) a concern with flexibility or robustness in the context of environmental change;

(iii) management, estimation (of cost and time), and control problems in the creation or installation of very large organizations or units of plant;

(iv) new problems of management and control arising in the operation of large organizations and units of plant;

(v) increased problems of statistical reliability or security which are accentuated by the concentration of capacity into fewer, larger centers.

A number of the research problems being undertaken by the Management and Technology Area at IIASA include some of the features above; particularly that on "Organization and Control." This paper aims to provide a broad review of the concepts and literature relevant to problems of scale. Its coverage is multi-disciplinary, and it represents a collaborative effort by scientists from differing socio-economic backgrounds. It is intended to serve as a point of reference in future research, and as a basis for discussion at future workshops concerned with "scale".

Abstract

This paper considers the general problem of how the "scale" or "size" of an entity is or should be determined, for entities ranging from individual units of plant to large organizations and industrial complexes. Several "levels" of scale are defined. The factors bearing on scale decisions are identified.

A number of techniques are reviewed, along with their relationship to socio-economic environment. In the socialist economies, the national and sectoral plans provide the framework for an analytical solution by mathematical programming, including non-linear production cost functions to represent economies of scale. In the market economies, the uncertainties of competition create a less stable environment; in which the scale decision has competitive significance. The relationship of large-scale projects to overall strategic planning is emphasized.

Following the review of techniques and methodology, the contribution of eight distinct disciplines to the subject is described. Section 5 considers research issues, discussing the problem of generalizing the measurement of scale, and emphasizing the changing nature of environments. Reference is made to the expanding East-West trade, and the growth of large-scale, long-term agreements. Within the Western economies, the pursuit of scale economies and of dominant market share may be leading to changes in the causal texture of operating environments.

In the final section, possible case material for future research is considered. The case of coal-fired electricity generating stations is reviewed. A description of the interaction between the growth of scale and of "relevant contexts", through diffusion and the reduction of barriers, leads towards consideration of possible implications in the field of trade planning and industrial development models. Other research problems in the field of industrial rationalization and restructuring are suggested.

TABLE OF CONTENTS

"PROBLEMS OF SCALE" - THE CASE FOR IIASA RESEARCH

1. INTRODUCTION - THE PURPOSE AND SCOPE OF THIS PAPER

1.1 Background: IIASA, MMT, Research Topics

IIASA - the International Institute for Applied Systems Analysis - is an international, but nongovernmental, research institution sponsored by scientific organizations from 17 nations, both East and West. It was established in October 1972 on the initiative of the United States and the Soviet Union to bring together scientists from different nations and different disciplines for joint investigation of problems of international importance, both global and universal in character.

o *Global problems* cut across national boundaries and cannot be resolved without the joint action of many nations. They include the problems arising from the need to satisfy mankind's needs for energy, food, and basic resources while protecting the global climate and environment.

o *Universal problems* lie within national boundaries, but are shared by all nations. They include the problems of providing adequate health care, transportation, housing, and other services to a nation's citizens, while preserving the national and regional environment.

The Institute's analyses are characterized by a focus on policy problems and a broad scope; they cut across traditional disciplinary, institutional, and national boundaries.

The origins and sponsorship of the Institute lead it to have three objectives:

o To promote international collaboration

o To advance science and systems analysis

o To apply its findings to problems of international importance.

Within IIASA, the "Areas" are the mechanism through which IIASA maintains contact with the boundaries of research in the large number of disciplines relevant to systems analysis. Of the four areas, one is "Management and Technology" (MMT).

The *Management and Technology* Area addresses issues arising from the ways in which societies design and manage organizations and technologies, and from their impacts on each other and the larger society. The disciplines of engineering, management science, information science, economics, and sociology (among others) are germane to these activities.

Within the area, 1978 sees the completion of a number of major projects, and the initial definition of a new research program for 1978-79. A basic principle is that the pursuit of methodological developments should arise out of the needs of real, current problems. Table 1 illustrates this interaction: the short-term program is organized around specific tasks or projects. The pursuit of these should contribute to the long-term objectives of methodological development on the research topics. This paper is about one of these research topics: "Problems of Scale."

Table 1: The Task/Topic Matrix

TASK/PROJECT	Program Management	Value and Risk	Innovation	Man/Computer Interaction	Problems of Scale	Management/ SA Interface
1. PROGRAM MANAGEMENT						
Shinkansen	X	X	X	X		
Health Program	X	X		X	X	X
2. ENVIRONMENTAL PROGRAM	X	X	X		X	X
3. MANAGEMENT OF TECHNOLOGY		X	X	X	X	X
4. USE OF MODELS IN POLICY FORMULATION		X	X	X		X
5. ORGANIZATION AND CONTROL	X	X		X	X	X
	LONG TERM OBJECTIVES					

SHORT TERM PROGRAM

1.2 Scope and Purpose of this Paper

This paper is intended as a "discussion document," designed to stimulate and advance a process of debate which will lead to continual review and amendment of the ideas and classifications presented. Its aim is to invite views and comments from scientists with relevant interests, and from planners and managers with responsibility for decisions in which scale is a significant parameter.

In the following sections, an attempt is made to classify and categorize the general subject of problems of scale. A brief description and summary is included of some of the principal concepts presented in the literature of various disciplines.

From this process of classification, review and comparison, we start to identify specific problems and shortcomings of current methods.

Finally we seek to define research objectives directed to overcoming these problems, and hence to the general development of systems analysis methodology.

2. BASIC DEFINITIONS OF TERMS AND IDEAS

2.1 "Problems of Scale" in Management Decision-Making

"Problems of Scale," as a title, is broad. Everything in the observable universe is in principle measurable, usually in many dimensions. "Problem" implies purpose, and we restrict attention to purposefully created artifacts and organizations; while not ignoring the possibilities of obtaining insights from natural systems (e.g. the evolution of species). The ultimate aim is to improve understanding, and therefore management capability, in certain broad classes of situation. These are situations where there is a choice between alternatives, and where a significant feature of the differences between the alternatives is their differences of scale.

Solving problems of scale is not a day to day or routine operational activity. For practising managers, the issue is related to medium or long-term planning, or to strategic rather than tactical management. A decision on scale is taken when one wants to establish or restructure an enterprise; to increase (by investment or purchasing) or decrease (by selling) the scope of an organization, often as part of a change of strategic policy or goals and objectives.

As is generally understood, the essentials of the management decision-making process* comprise the following four principal stages:

---

*For example, as described by authors such as:
Green, P.E., Tull, D.S., Research for marketing decisions. Englewood Cliffs (N.J.), 1966, p.64; Horngren, D.T., Cost accounting: a managerial emphasis. Englewood Cliffs (N.J), 1967, p.777; Richards, M.D., Greenlow, P.S., Management decision-making. Homewood (Ill.), 1966, p.53; Emery W., Niland, P., Making management decisions. Boston, 1968, p.9; Kepner Ch. H., Tregoe B.B., The rational Manager. N.Y., 1965, p.179; Morris W.T., Management science (A Bayesian introduction). Englewood Cliffs (N.J.), 1968, p.6; Elton S., What is a decision? "Management Sci.", 1968, N. 4, p. B-173; Cleland D.T., King W.R., Management: a system approach. N.Y. 1972, p.226; Drucker P.F., How to make business decision. - In: Decision and information systems. W.T. Greenwood (Ed.). Cincinnati (Ohio). 1968, p. 53; McGrimmon K.R. Managerial decision making. - In: Contemporary management (issues and viewpoints); J.W. McGuire (Ed.). Englewood Cliffs (N.J.) 1974, p.445.
Also in: *Amerikanskii kapitalizm i upravlencheskie reshenija* (American Capitalism and management solutions), Nauka, Moscow, 1976.

1. Setting of objectives of organizational activity.

2. Identification and analysis of problem.

3. Generation of alternative courses of action and analyses of probable consequences.

4. Choice of alternative and detailed evaluation.

Clearly the problem of scale cannot exist in a vacuum. The problem is set within existing goals and objectives; physical location; management culture of the organization; etc. Therefore a "Problem of Scale" cannot be interpreted as one of strategy, or as a problem of goal-setting. It is not related to Stage 1 of management decision-making, but a little to Stage 2 and much to Stages 3 and 4.

Within the socialist countries, the centralized planning system provides guide-lines and objectives for regional and industrial planning. This is stage 1 of the above four. The problem of determining scale is therefore equivalent to the problem of location and planning of production facilities to meet in the most efficient way the objectives of the plans. Within a company in a market economy, the objectives are less controlled, and the decisions on scale may therefore interact with consideration of objectives. But in general, questions such as how large a hospital, a colliery, or an enterprise should be arise when one starts to generate alternative courses, analyse probable consequences, choose alternatives and evaluate them.

Thus the problem of scale is one of alternatives, but not a problem of goal-setting. To take scale into account in the processes of generation of alternative courses and analysis of probable consequences (Stage 3), and of choice and evaluation of alternatives (Stage 4), one needs to know the factors of scale and to have criteria.

The generator of alternatives takes into account all factors of scale (political, social, economic, organizational, etc. - see 2.4 below) and uses many possible criteria in the analyses of probable consequences. But the evaluator uses only some criteria, which are crucial from his point of view (e.g. operating efficiency, organizational complexity, flexibility, risk, social consequences, security). The criteria defined as crucial depend on many things. As an example, flexibility of a corporation depends not only on scale, but on organizational structure, management system, etc. One needs to add that the set of crucial criteria depends on people, their experience, and the environment in which the choice of alternative takes place. For instance in countries with centralized planning, the risk of bankruptcy does not exist.

It should be mentioned that criteria for evaluation of alternatives include both those of organizational effectiveness (as influenced by the process of organizational design), and those performance characteristics arising in the operating process. Therefore a distinction should be drawn between the study of problems of scale, which is one topic; the study of organizational effectiveness and/or efficiency, which is a second topic; and the influence of scale on organizational effectiveness, which is a third topic. This paper seeks to clarify the definition of the problem we want to study.

Having reviewed the place of scale in management decision-making one might ask: what kind of research on problems on scale could be launched in IIASA - academic, or applied? Should the eventual result be in the nature of a text-book, or a hand-book?

If one takes into account the fact that questions of scale are only meaningful within their relevant context, of strategies, goals, objectives, location of organization, etc., the answer can only be a text-book. For a hand-book should give specific instructions, related to a specific relevant context; if we concentrate on general principles and general methods, then the result must be more like a text-book.

2.2 Levels of Scale

A useful sub-division of problems of scale is the distinction between the following "levels": the terms underlined will be used in this sense in the remainder of the paper.

Level 1(a): the scale of a single unit of physical equipment: the "engineering level" or "unit level"

(b): the scale of a single product line (which might be produced by several separate units of equipment)

Level 2: the scale of a single plant or factory (i.e. on, or based on, one site; but possibly containing several engineering units or product lines): the "plant level"

Levels one and two coincide in the case of a single-unit (or "single-train") plant, which typically depends on a single major component.

Level 3: the scale of a single organization: the "corporate level" or "organization level"

Level 3 is less clearly definable in operationally unambiguous ways, and in terms capable of clear and standard interpretation in different countries. For instance, it may coincide with level 2 in a single-factory company. In a company comprising several plants engaged in similar activities, the plants might collectively be viewed as a single organization; but this company might itself be a subsidiary part of a larger company. This membership of a larger unit could be relevant to financial and negotiating strength, and therefore in wider dimensions as a result; but might be irrelevant to the company's technical efficiency. "Organizational level" thus requires careful definition, particularly where comparisons are being made: a "big" organization could be "small" in the scale of its activities in a specific field.

Level 4: the scale of national economic programs and industrial complexes: "co-operative level"

During recent decades, new organizational forms of large-scale national economic programs (for example TVA in the USA [32]) and territorial/industrial complexes (for example Bratsk-Ilimsk territorial production complex in the USSR [46]) have come into being both in the Western and in the Eastern countries. In the MMT Research Plan 1978-79 of the Institute, a "program" is defined as "the process of implementing a decision to create change. There is usually a limited set of objectives; thus the program lasts for only a given period of time. Normally, it is organized on an ad hoc basis, lying outside the continuing bureaucratic machinery. In general, a governmental decision is involved."

An industrial complex can be defined as a set of industrial enterprises, located, in order to raise efficiency, on one site or in neighbouring geographical locations, and having a common infrastructure. For a clear description of this new entity, the industrial complex, and for an explanation of the efficiency of its establishment, we reproduce a description from a USSR source [7]:

> "Depending on the nature of the enterprises they contain, industrial complexes may be divided into three groups, as follows: (1) those comprising heterogeneous, unlinked enterprises; (2) those comprising enterprises that are allied technologically; and (3) those comprising both the preceding groups.
>
> Heterogeneous enterprises situated in one geographical location may have a common power system, a single system of water supply, sewerage, water purification, and other engineering services and communications. The setting-up of an integrated system of transport and warehousing facilities also produces great benefits.
>
> Thus, the length of railway lines within the area of an industrial complex can be reduced by 18 to 47 per cent and of roads by 9 to 30 per cent. The establishment of an integrated system of servicing and ancillary enterprises results in substantial savings in capital investment and operating costs, and enables rational use to be made of electricity, fuel, and water. A reduction of 20 to 40 per cent in the ground space occupied by industrial enterprises is also of no little significance.

The creation of industrial complexes of the second and third groups provides incomparably greater benefits.

When technologically allied enterprises are grouped together in one location, the savings obtained through cooperation of ancillary and preparatory industries and stockpiles, are added to the advantages mentioned above.

Savings can be made in capital investments by reducing the production area occupied by ancillary and preparatory shops by 25 to 40 per cent, and by reducing the amount of equipment by 35 to 50 per cent. Operating costs are also reduced.

Finally, savings are made by coordinating the use of raw materials and supplies by several enterprises or by combining their consecutive technological processing at various stages."

"The creation of industrial complexes must be closely tied up with the development of a rational system of towns and a uniform settlement policy for the country. It encourages the establishment of common, joint construction facilities, saving 20 to 40 per cent on capital investment, and united residential areas meeting the requirements of science and the technological possibilities of the building industry. Solution of all these problems necessitates close cooperation between the regional planning agencies and sectoral and town planning and building institutes."

The characteristic features of the Soviet Union's individual regions will be increasingly determined by the implementation of major economic programs and the establishment of territorial production/industrial complexes [50]. As examples one could name the program of development of agriculture of the Non-Black Earth Zone (a region of low fertility), the program of development of industrial-agrarian zone, of the Kursk magnetic anomaly, etc.

The USSR is paying great attention to the establishment of these territorial-production complexes, as they are considered the most efficient direction of economic development in the conditions of centralized planning. The West Siberian territorial-industrial complex, the system of Angara-Yenisei complexes, the South Tajik complex, and others have already been launched.

The formation of the new Timano-Pechora industrial complex, with the use of the large oil and gas deposits in the area, will get off the ground; and in the long term the USSR will launch a number of complexes gravitating towards the Baikal-Amur Railway now under construction. The creation of such complexes raises new managerial problems, such as for example the appropriate organizational forms of co-ordinating the planning, construction, operation and development of large-scale complexes.

In conclusion one could say that the emergence of new levels of scale raises new problems. Large-scale major economic programs (Alaska in the USA, the Non-Black Earth Zone in the USSR, and industrial complexes (the Invergordon chemicals complex in Scotland)) require new forms of management: cooperative management, e.g. joint management of some corporation, companies, or industries. The methodological problems of cooperative management of large-scale programs and complexes is a major topic of study in IIASA.

A development similar in some respects to the industrial complex is that of very large scale joint ventures, often involving international agreements. Their creation requires co-operative management, similar to the level 4 defined above; but once created, they can become essentially unified organizations, similar in their characteristics to level 3.

2.3 The Scale of Environment, or "Relevant Context"

The word "environment" is commonly used, but with a very general meaning. In studying an economic or industrial entity at any level, the systems analyst views it as part of a "system": a "set of interrelated elements, each of which is related directly or indirectly to every other element, and no subset of which is unrelated to any other subset" (Ackoff [2])*. The entities described by Levels 1 to 4 above are not complete "systems" for the purposes of our study, because the questions raised by scale alternatives have to take account of relationships with the "environment." We use below the term "relevant context" for those parts of the general environment which are relevant to the determination of appropriate scale in a particular case; in other words, the "system" to be studied is the entity (machine, factory, organization) and its relevant context.

---

*A fuller definition is given by Allport [4]:

> "... any recognizable delimited aggregate of dynamic elements that are in some way interconnected and interdependent and that continue to operate together according to certain laws and in such a way as to produce some characteristic total effect. A system, in other words, is something that is concerned with some kind of activity and preserves a kind of integration and unity; and a particular system can be recognized as distinct from other systems to which, however, it may be dynamically related. Systems may be complex, they may be made up of inter-dependent sub-systems, each of which, though less autonomous than the entire aggregate, is nevertheless fairly distinguishable in operation."

Defining the boundary of the relevant context is sometimes difficult, but always important. It is important, because (as discussed further in 5.2), the measurement of scale at levels 1 to 4 is virtually inseparable from the definition of the scale of the relevant context. A hospital might be "too large" to serve the local town, but "too small" to serve the surrounding region; which context is relevant? The relevant context can take many forms, such as the following:

"everywhere within 200 kms. of the plant"
"all owners of VW cars"
"the whole of the industry"
"the whole market"
"the national economy"
"Eastern Canada"
"the Comecon countries"
"the world"

For example, a statement such as, "this country is too small to justify a car industry, but might consider an assembly plant" is full of implications and assumptions on all levels of scale, as well as the scale of relevant context.

Because of the interactions between the different levels, it is common to find different descriptions of similar problems: for instance, within a country, the "location of productive facilities" to meet the country's needs implies a decision also on the "scale of production" in each plant.

The same entity may be viewed on different levels in different contexts. A seaport's capacity might be a level 1 scale problem in the context of national strategy; but it is level 2 when we consider the design and scale of the individual docks. A country might be the relevant context for some industries, but a "level 3" organization in relation to supra-national negotiations about trading areas.

No commitment has yet been made as to which levels are to be the subject of study: this is a question we return to in section 6.

2.4 Factors of Scale, Static and Dynamic

All determinants or factors affecting the choice of scale for an entity can be grouped in different ways, depending on the concrete situation and the research goal. For example, the factors might be grouped as follows:

- political
- social
- economic
- technological

- organizational
- managerial
- financial

Each group of factors could be further subdivided. As an example of a political factor one can cite security. Another political example could be the desire to create "the largest (smallest, longest, etc.) in the world," which might be established to surprise and impress the world.

Social factors such as the problem of employment/unemployment in a certain town or region could be of crucial practical importance in determining the scale of a business enterprise.

The political and social factors require in many cases the creation in practice of entities on a scale which is far from optimal on economic grounds. The role of political and social factors becomes crucial only in the solution of practical problems of entity scale in a definite location or region, but they could not affect the general determination of optimal scale. Therefore such groups of factors cannot so readily be generalized and taken into account in our research, although we must recognize their existence.

All other factors are generally significant in determining optimal scale. Differing factors influence the scale of organization or of its units, in opposite directions: some of them favouring an increase of scale, some a decrease. A general feature to be observed is that factors favouring the increase of scale are mainly internal, while those which favour decrease of scale are mainly external.

For instance, in manufacturing industry, we have the following set of internal and external factors affecting the scale of plant:

Table 2: Factors affecting the scale of plant

Increase (mainly internal):	Decrease (mainly external):
- equipment	- economic and geographical circumstances of distribution
- technology	- location of consumption of goods
- organization and management of production	
?	?
?	?

The balance between internal and external factors determines the static scale of entity. Quantitative analytical techniques, taking account of both increasing and decreasing (internal and external) factors, can be used to determine "optimal" scale. All these factors (both internal and external) can be considered as direct factors determining the scale of entity in a static framework.

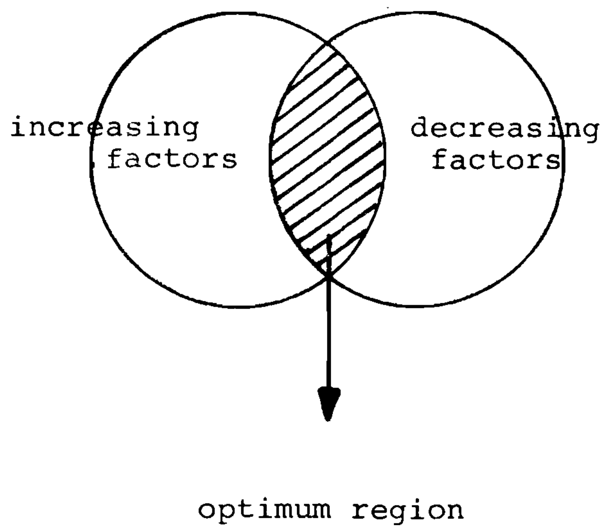

Figure 1: Static influences on the scale of organization

Historically we know that the scale of enterprises has not been of constant magnitude, but has generally had an increasing trend. To understand this phenomenon one must add to the list of factors one more: time. Each period of development is characterized by certain levels of development of machines, technology, organization, management, forms of production, economic and geographical conditions, etc. [73]. In other words, the factors and their weights are changing over time. But the rate of change of differing factors are not the same. Internal factors, which are determined by scientific progress and technological changes, are more dynamic, and external factors are changing less quickly. Therefore the scale has tended to increase. Thus the time factor changes the action of direct factors on scale. But this influence is carried out in an indirect way through change of concentration, specialization and cooperative forms of production, change of equipment and its composition, of consumers and their needs, of conditions of transportation and of business connections, etc. As a result, one can add to the classification of factors of scale one more group: indirect factors.

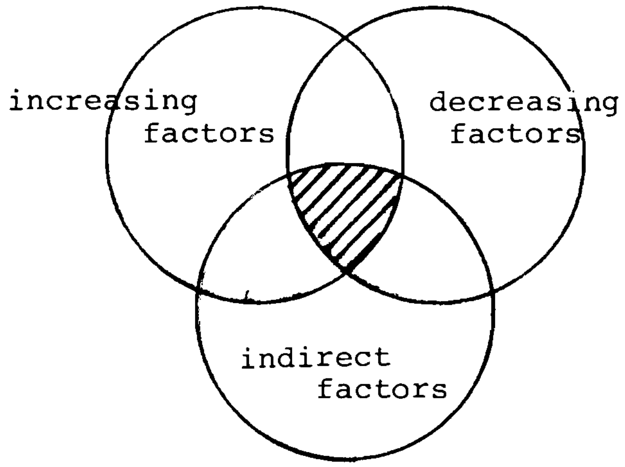

Figure 2: Static and Dynamic influences on the scale of organization

The impact of different economic factors on problems of scale is represented graphically by Figure 3. This example is related mainly to the plant in the processing industries, but it takes place in each problem of scale.

Figure 3 shows how in manufacturing industry, scale of enterprises depends on many economic factors both direct and indirect. Among them are technology, organization and management, transportation, mineral resources and materials, production needs, division of labour, specialization, cooperation and concentration forms and so on. If one takes into account other groups of factors (political, social, environmental) the set of scale factors will be very large. Each industry or organization has its own technology, organizational peculiarities, particular locations, distribution pattern, goals and objectives, managerial cultures, customers and so on. Therefore one can put the questions:

1. Could there be elaborated a general methodology of scale for determining the size of hospital, super tanker, agricultural farm, industrial plant, research and development organization, and so on? In other words whether one could generalize factors of scale, i.e., the technologies, organizational peculiarities, regional peculiarities, policies and strategies, goals, and objectives, environments, and so on of different industries, types of organization, and states? How would one find common elements of entity scale?

2. Could there exist the general problem of entity scale, or must there exist problems of scale peculiar to (a) each industry (mining, processing, agriculture and so on) or subindustry (coal in mining, machine tools in processing, health and education in service industry), (b) each level of scale? What purpose would such generalization serve?

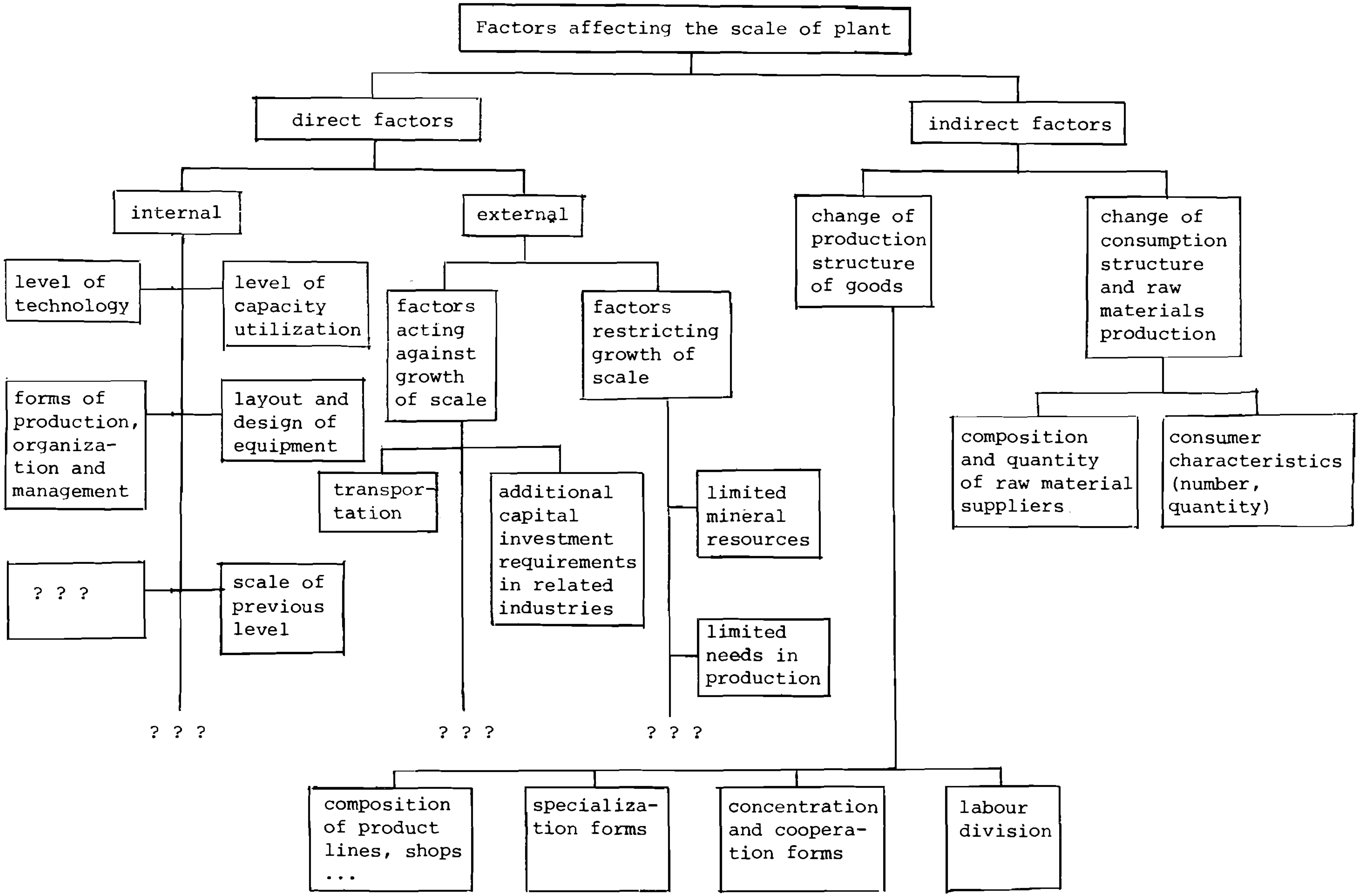

Figure 3: Economic factors influencing scale

2.5 Relations between Factors and levels of Scale

We suppose that there may exist relations between the factors and the levels of scale, and that scale on each level is related to a definite set of factors. For example, level 1 (unit level) is affected more by technological factors than by political, social or managerial.

Levels/factors	techno-logical	organi-zational	mana-gerial	finan-cial	eco-nomic	social	polit-ical
1. (Unit, product line) 2. (Factory) 3. (Organization) 4. (Economic program and Industrial Complex)							

Figure 4: Influence of factors on levels of scale

It is our opinion that investigation of the relations between factors and levels could help us to answer the question: How large should be, say, an organization?

2.6 Scale and production homogeneity/heterogeneity

Literature study shows many examples in which scale increase provides performance improvement. As an example we quote these performance characteristics from Soviet Union source [9]:

Table 3: Performance characteristics of thermal power-stations

	Thermal power-station capacity (MW)				
Performance characteristics	200	300	600	1200	2400
Specific capital investment	100	86	75	66	60
Specific volume of main building	100	88	84	58	51
Construction and installation as % of total capital cost	66.5	64.0	60.0	50.5	45.0
Specific quantity of operating personnel	100	84	60	32	24
Electrical energy production cost (price 10 rub. per 1t. of eq. fuel)	100	91	87	78	70

The same phenomenon takes place in electrical energy, iron and steel production, many branches of chemical industry, mining (coal, iron-ore) industry or in one product industries using thermal energetic, chemical and mining technologies. In these industries, scale and performance characteristics are strongly dependent on capacity and productivity of machines/equipment. In this group of industries the differences in scale among small, medium and large lie mainly in the capacity and productivity of equipment - turbogenerators, boilers, blast-furnaces, etc.

Therefore the enterprises in these industries use universal technologies, and seek a high level of utilization of capacity. We label this group of industries as "technology homogeneous" industries. Efficiency characteristics in these industries depend as much on specific capital investments, as on specific norms of material and fuel expenditures, both of which are smaller when capacity of equipment is higher. The result of scale growth in technology homogeneous industries is reduction of both capital and operating cost. One could suppose and stress that the cause of improvement in the performance characteristics with scale growth is a definite homogeneity of production.

But practice shows that scale increase sometimes does not provide improvement of performance characteristics. One could give many examples from different countries, where small enterprises co-exist on an equal footing with medium and large ones. This phenomenon takes place in machine tools, electrotechnical, radiotechnical, electronic, textile, meat production, etc.: in other words in industries which are based on the use of mechanical-technological methods, and local technologies. In these industries scale depends mainly on quantity of homogeneous equipment, but not on the capacity and productivity. Therefore the efficiency, i.e. ratio of resource input to output, in small and large scale units is approximately the same; and small scale can exist in a competitive environment. In these industries, scale increase can sometimes cause decrease of performance characteristics, and many countries could give examples of bankruptcy resulting from misunderstanding of the relation between scale and homogeneity of production. The improvement of performance characteristics in these industries depends not so much on scale increase, as on improvement of local production methods, organization, and management of production.

Study of both Eastern [47] and Western sources shows that many countries have established systematic procedures for achieving a greater degree of homogeneity. All these ways can be divided into two groups:

(a) organizational or economic ways to enable the concentration of homogeneous products or homogeneous technological processes in a single place. This leads to specialization of production.

(b) engineering ways of achieving increased homogeneity of products, technological processes, operations, etc., or in other words standardization of production.

Industrially developed countries are paying great attention to the problem of achieving a greater degree of homogeneity of production, and consequent performance improvement. One of us, in a study of the problem of labour mechanization and automation in the Lithuanian SSR [31] and USSR [34] industries, concluded that optimal scale of homogeneous production, together with development of specialization and standardization, forms the basis for efficiency increase in the auxiliary production sectors of the Soviet Union industry.

Both Eastern and Western countries' experience show that specialization and standardization are playing a crucial role in the improvement of performance characteristics. As a further example, one could note the use in the U.S.A. of the broad application system "Simplification-standardization-specialization" and in the USSR of the system "Standardization-specialization-automation."

2.7 Optimising the Scale: Minimum, Maximum, Mix: Which Problems?

Much of the literature on scale is apparently concerned with the determination of the "optimum size" of an entity on one of the four levels. "Optimum" implies the reduction of all criteria to a single dimension on which alternatives can be ranked, e.g. "cost" or "efficiency." "Optimum size" implies that the only, or the major significant, difference between the alternatives is that of size. The picture is often summarized as in Figure 5, with its assumption of the single optimum size, and monotonic worsening of performance the further the size deviates from this, above or below.

Few real situations are as simple as this, for reasons including the following.

1. There is not a universally agreed, single measure of "good" and "bad": the performance measures are multi-dimensional.

2. There is not a single point in time at which a decision must be made, but a succession of decisions, which will influence the alternatives available at subsequent times.

3. The evaluation procedure depends on data (especially forecasts) which are unknown or uncertain, unless they are provided by higher level plans.

4. The evaluation procedure requires assumptions about causal connections which are not fully understood.

5. The alternatives available may be a finite number rather than a continuous range.

6. There may be several independent parties interested in, and affected by, the decision; so that the final decision may be a matter for negotiation or arbitration between these interests.

These conditions restrict the applicability of simple optimizing techniques in many problems of scale.

The phrase "minimum viable size" is another term often used in discussions of scale problems, reflecting an assumption such as that shown in Figure 6: the L-shaped cost-curve. This suggests a "Level 1" viewpoint: that the inherent nature of the product or process, e.g. for engineering reasons, makes it prohibitively expensive to contemplate very small scale. This may be correct. However, such curves are sometimes based on a single technology and pattern of organization, appropriate to a certain size; and if one really wanted a small-scale unit, some cheaper approach might be found. Gold [36] emphasized the close relationship, for instance, between scale and specialization of function.

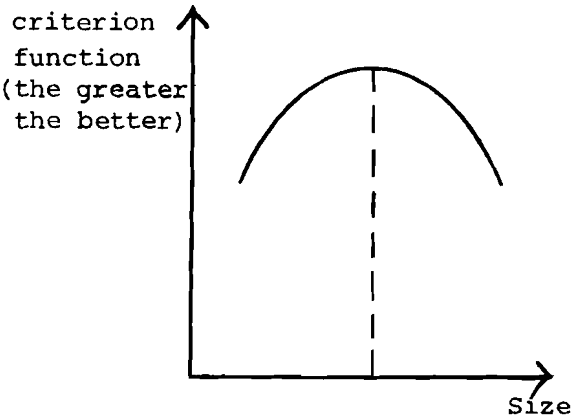

Figure 5: "Optimum Size"

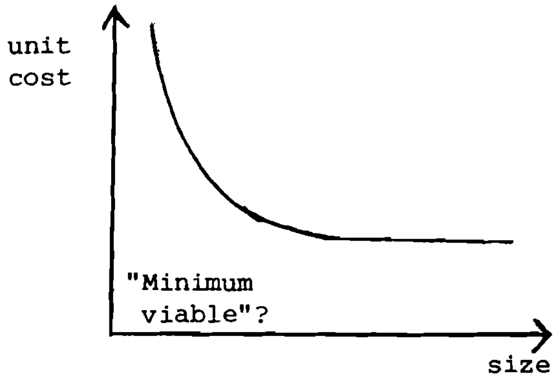

Figure 6: "L-shaped" Cost Curve

Another important aspect relating to "minimum viable size" is the consideration of "relevant context." In a competitive environment, the minimum viable size may be determined, via Figure 2, from the lowest cost achievable by competing organizations: this may have to be matched. But transport costs, tariff or quota barriers, product differentiation and many other factors could alter the situation. Thus "minimum viable size" is a question whose resolution requires consideration of the environment as well as of the unit itself.

"Maximum scale" may similarly be determined by engineering limits and/or the local natural environment; or by limits on level 2 (factory space) or level 3 (organizational - e.g. financial limits, or problems of organizational complexity); or by relevant context - e.g. total market potential, or regional plan requirements.

In many situations, the diversity of the environment calls for a diversity of responses, and there is no single "optimum size." For example, in transport, a transport organization may need both large vehicles (for low cost bulk haulage) and small ones (for small local deliveries). In such situations, the relevant problem is that of determining an "optimum mix" of units of different size. In many industries, there are good reasons, and not only historical ones, for the co-existence of plants of different sizes, and of organizations of different sizes.

To show the complexity of optimising the scale of an entity, we use as example level 4, the industrial complex of inter-organizational level. Optimising the industrial complex's scale requires clarification of the links with a number of other problems. The scale of a complex depends on:

- natural and geographical conditions, location
- sectoral structure of the enterprises in the complex
- technological and economic links of enterprises within the complex, region or industry
- degree of production concentration on the enterprises of complex
- availability/unavailability of labour, utilization of available labour
- optimization of the enterprises' sizes, taking into account their specialization/combination

Several criteria could be used to establish the optimum scale of an industrial complex. One of these, for example used in the USSR, is to ensure the maximum increase in the efficiency of production, given the volume of production planned for the economic region [55]. A number of Soviet authors give fuller details of the planning methods employed in these optimization studies, and further details are given in the appropriate sections below, on techniques.

This brief introduction to some of the terms commonly used in describing problems of scale indicates their potential diversity of form and content. To some classes of problem, there exist "methods of solution," and where satisfactory solutions exist for clearly defined problems, there is no need for research. The research need and interest will be greatest where changing circumstances are creating new problems, not yet fully understood or well-defined, and to which existing methods do not provide adequate solutions. It is the purpose of this paper to identify such classes of problems.

3. TECHNIQUES, MODELS, METHODS AND METHODOLOGICAL DEVELOPMENT

3.1 Introduction: Differences of Environment

The basis of generalization about problems or alternatives of scale must be the identification of common features in superficially different situations, and the development of conclusions systematically related to those features; so that they can be applied to any other situation in which these features are present.

It is clear that there are some fundamental differences between the economic systems of East and West, and these affect the techniques appropriate to the consideration of problems of scale in two ways:

(a) the environment of the management decision-maker

(b) the general goals and objectives of economic development.

Socialist planning in the East, and economic programming (indicative planning) in the West, differ as regards their social character, principles and functions, owing to the basic difference between property relations in conditions of social and of private ownership.

Both socialist planning and indicative planning have to solve one of the most important problems of economic development: economic efficiency [7]. But the principles used in solving these problems are fundamentally different. In conditions of social ownership, the basic criterion of the efficiency of social production is achievement of the best results at the least cost in the interest of society. Private ownership does not approach the problem of maximum satisfaction of social needs on the basis of efficient use of society's resources, as there can exist basic conflict between efficiency from the point of view of the total society, and the objectives of individual persons or groups of persons joined together in a corporate organization.

A socialist economy is based on a set of interconnected and coordinated plans, main among which are macro-economic (i.e. national), sectoral (i.e. ministerial), regional, and enterprise plans. The national plan is the central element of socialist planning, and is based on the sectoral and regional plans; while the latter in turn rest on enterprise plans. The centralized macro economic planning carried out in the Soviet Union and other socialist countries is directive planning. The economic programming carried out by governments in the West is only indicative planning, and generally limited to making recommendations to the private sector, which may or may not be implemented, depending on the interests of the private sector, and on the objectives and strategic plans of private corporations.

From the description above one can conclude that the environment in the socialist countries is determined by the system of plans. These are described as follows in Volume 1 of "Planning of Socialist Economy" [7]:

> "USSR Gosplan draws up the state plan for the development of the whole economy and fixes the assignments for USSR ministries and departments and for Union republics in the form of aggregate indicators. USSR ministries and departments compile more detailed centralised plans for the development of particular economic sectors and industries. The Gosplans of Union republics compile plans for the complex development of the republic's economy as a whole and plans for the areas of economic activity under republican control which contain assignments for the various ministries, autonomous republics, regions, etc., and serve as the starting point for the drafting of similar plans by Union republican ministries, the Gosplans and ministries of autonomous republics, and the planning commissions, boards, and departments of local authorities.
>
> Thus, the following planning system operates at the present time:
>
> (1) the state plan for the development of the Soviet economy;
>
> (2) plans drawn up by USSR ministries and departments for their economic sectors and industries;
>
> (3) plans compiled in the Union republic regarding the republic's economy as a whole and those areas of economic activity under republican control;
>
> (4) plans drawn up by Union republican ministries and departments for the sectors for which they are responsible;
>
> (5) the plans of autonomous republics regarding their economy as a whole and those areas of economic activity directly in their control;
>
> (6) the plans drawn up by the ministries and departments of autonomous republics for their own sectors;
>
> (7) the economic plans of territories, regions, towns, etc.;
>
> (8) the plans drawn up by boards (departments) of local executive committees;

(9) the plans of amalgamated enterprises, single enterprises, organisations, and institutions.

All these plans--the united state plan of the USSR, the plans compiled by the USSR ministries and departments, the plans made by Union and autonomous republics and all the other areas of the economy--are closely interconnected and form a single system. This single planning system ensures centralised planned management of the economy and the development of initiative on the part of local bodies together with the economic independence of enterprises."

It is necessary to say that in solving the problems of economic development including problems of scale, socialist planning makes it possible to ahcieve coordination among all those participating in production, and bewteen the interests of the whole economy and its various branches, economic regions and enterprises. The central macro-economic plan fully takes into account both social needs and economic resources throughout the country.

In the socialist countries, alternatives of scale form part of the more general problems of planning the location of industries, which are an integral part of macro-economic planning. Such planning has to take account of the specific character of each sector, its technological and technical features, the nature of its raw material base, the consumption of materials, transportability of its product, etc., as well as the natural and geographical features and economic resources of each economic area. Methodological principles of planning the location of the country's productive forces including scaling of production over the long term and specific order are worked out in the Soviet Union as follows [7]:

"... The first stage consists in drawing up <u>sectoral</u> schemes for the development and location of <u>industry</u>, and the second stage in drawing up schemes for its development and location in <u>economic areas</u> and <u>republics</u>. The third stage comprises the <u>compilation</u> of a <u>General Scheme for the location of productive forces</u> in the <u>Soviet Union, which coordinates and resolves</u> any inconsistencies between the sectoral and regional schemes. These schemes provide the basis for formulating the requirements as to the location of industry in sectoral, regional, and macro-economic plans. This order of operations makes it possible to integrate sectoral planning with regional planning, taking into consideration the development interests both of sectors and economic regions."

The principles for siting and scaling enterprises have to be subdivided into general principles, i.e. applicable to all sectors of production, and specific, i.e. sectoral ones, applicable to separate sectors.

The designation of general principles is to ensure maximum economic efficiency of production in a sector, and of sectoral principles to ensure the necessary volume of output with the minimum possible expenditure of labour and other resources.

In the market economies of the West, the environment of the corporation is less certain; firstly because of uncertainty about growth of the total market, and secondly because of uncertainty about the behaviour of competitors. This means a greater degree of risk is attached to very large-scale, long-term commitments; but at the same time, competitive pressures encourage firms to seek maximum economy of scale, and can lead to the elimination of smaller scale producers unless their position is deliberately defended by government action (e.g. tariffs, quotas, special grants).

It is important to emphasize these differences of environment, because they lead to the use of quite different techniques in the study of alternatives of scale. Some of these techniques are described below, and the same issues recur in later sections on the measurement of size (5.2) and on the modelling of environment (5.3).

The most basic numerical techniques of evaluation are mathematical, and some of these are introduced below (3.2). At a more general level, we consider the development of "standard models" (3.3) and "standard methods" (3.4), and then in 3.5, we extend the discussion to consider directions for the development of methodology.

3.2 Mathematical Techniques

Pure mathematics is devoid of "content," serving merely as a language any discipline may use to express quantitatively its concepts, measurements and relationships. But within such general quantitative disciplines as operational research, certain models and techniques have been found to have widespread applicability. These tools have been developed and improved through such practical use, and through the parallel development of improved computational facilities. Three are reviewed here as being potentially applicable to the consideration of scale alternatives.

3.2.1 Mathematical Programming.

Both in the Soviet Union [41 and 45] and in the United States, extensive theoretical development has taken place since the 1930s and 1940s in the problem of optimizing a single linear objective function under linear constraints. This is a technique of widespread use, wherever one has a clearly defined objective and constraints which can (at least to a satisfactory approximation) be represented in this form.

Its role within the planning system of the socialist countries is illustrated by the following quotation from the book edited by L. Ya. Berry. The sectoral planning is solved by linear programming. The economies of scale in production create a non-linear problem, (i.e. unit costs depend on quantity produced) and this requires more sophisticated techniques of mathematical programming. These are still quite manageable with modern computers and software, and are used in Stage 2 of sectoral planning to solve the questions of scale.

> "*Formulation of the problem:* to determine optimum structure of production in the given sector in each region, provided that the costs of transporting the products, given the existing (known) sources of raw materials, will be minimal.

Notation Used in the Model

Known quantities

- n--the number of regions, each of which we will denote by i, where $i = 1, \ldots, n$;
- l--the number of types of available output, each of which we will denote by j where $l = 1, \ldots, l$;
- P_i--the quantity of raw materials in the ith region;
- λ_j--standard consumption of raw materials on the jth product;
- Q_{ij}--annual demand for the jth product in the ith region;
- a_{jrd}--cost of transporting a unit of the jth product from the surplus region (r) to the deficit region (d) at current freight rates;
- c_{ji}--unit operating costs of the production of a unit of the jth product in the ith region;
- U_j--sectoral average of unit capital costs for the production of the jth product;
- h_{ij}--regional coefficient for adjusting unit capital investments in the jth product in the ith region;
- E--standard coefficient for the efficiency of capital investment.

Sought-for quantities

- x_{ji}--volume of production of the jth product in the ith region;
- x_{jrd}--quantity of the jth product transported from the rth region to the dth region.
- $x_{ji} \geqslant 0$--volume of output cannot be a negative quantity; this also applies to the volume of freight $x_{jrd} \geqslant 0$;

$\sum_j x_{ji}\lambda_j \leqslant P_i$--consumption of raw materials in the production of all types of output in the ith region must not be greater than the raw material resources in that region;

$\sum_r x_{jrd} = Q_{dj}$--total volume of the jth product brought into region d from other regions equals the demand in region d for that product;

$\sum_d x_{jrd} = x_{jr}$--total volume of the jth product transported from the rth region equals the volume of production of the jth product in that region.

It is required to determine the minimum of the functional:

$$\min\{\sum_{j,r,d} a_{jrd}x_{jrd} + \sum_{j,i} c_{ji}x_{ji} + E\sum_{j,i} h_{ij}U_j x_{ji}\}.$$

When Stage One of the problem has been completed, the optimum structure for the production of the sector's goods in each region is known.

Stage Two. Now that it is known which products each economic area (or region) must produce, and in what quantities, we turn to the question of determining the actual sites for enterprises, and their capacities. Thus, for example, if, as a result of solving Stage One, it is found that it is necessary to increase production of one of the products in an economic area, the best possible variant must be found for the location of the enterprises to be built, i.e. it is a question of formulating the problem of the location of enterprises producing homogeneous output in relation to the sources of raw materials.

In solving this problem, the main factors to be taken into consideration are manufacturing costs and transport outlays, and these vary inversely. Manufacturing costs depend to a great extent on the volume of production. When the volume of production within an enterprise increases, unit costs decrease. At the same time, however, other things being equal, the radius within which raw materials are transported is extended, with a consequent rise in transport costs. The effect of both these factors, therefore, must be taken into account, and a combination of the two should be found that offers minimum total costs.

In determining the optimum scale of enterprises, and deciding on their location, it is important to increase the efficiency of capital expenditure, i.e. to reduce the amount of capital expenditure per unit of fixed capacity, and cut down the recoupment period.

Unit capital investment and recoupment periods consequently must also be taken into account as well as the costs mentioned above.

The range of initial data depends on the nature of the problem and its formulation. When it concerns the location of enterprises in relation to sources of raw materials, the basic initial data are as follows:

(i) the annual resources of raw materials within the area selected for this particular problem;

(ii) the location of sources of marketable raw materials and the amount of raw material available from them;

(iii) the productive capacities of existing enterprises, and the additional capacity that must be obtained by building new enterprises and extending existing ones so that all available marketable supplies of raw materials will be fully processed;

(v) the cost of processing a unit of raw material, in accordance with the volume of production; when the capacities of the new plants are greater than any one of those already operational, processing costs are determined approximately, taking the indicators of standard designs in two variants: (a) when the plant is operating at full capacity, and (b) when it is operating under capacity;

(vi) the unit capital investment required for building new enterprises and reconstructing others that are already operational, in accordance with plant capacity and local conditions.

The initial data must satisfy several general requirements. Above all, the statistics must be sufficiently large and, as far as possible, evenly distributed over a period of time within which the effects of various features on an economic indicator can be examined, which makes it possible to determine the correlation between them with the necessary degree of accuracy.

Unknowns are the scale on which raw material must be processed at each possible site, and the scale on which raw material will be transported from source to processing enterprises.

In solving the problem, one has to try and find the location variant and the volume of production at each enterprise that will minimise the total outlay on processing all supplies of the raw material, transporting raw material to the processing point, and on capital investment, taking account of the recoupment period. This is, in fact, the optimality criterion in this instance.

Thus, we have:

p--number of points of raw-material production, each of which we will denote by r, where $r = 1, \ldots, p$;
n--number of possible points for producing the finished product, each of which we will denote by i (i.e. $i = 1, \ldots,$).

Sought-for Quantities

x_i--volume of production of output at the ith point; volume of production will be measured by the quantity of raw material processed at a given point;
Z_{ri}--quantity of raw material transported from the rth source of raw material to the ith point of production of completed output.

Known quantities

Q_r--quantity of raw material produced at the rth point;
$g_i(x_i)$--cost of production per unit of finished product at the ith point of production, depending upon the scale of production at that point;
$U_i(x_i - x_i^0)$--unit capital expenditure at the ith point of production, depending upon an increase in capacity above that of already operational enterprises (if $x_i > x_i^0$), or depending upon the scale of production at the new installation (if $x_i^0 = 0$);
a_{ri}--cost of transporting one unit of raw material from the rth point of raw material production to the ith processing point.

1. The quantity of processed raw material must equal the amount of raw material received from all sources;

$$x_i = \sum_r Z_{ri} \quad .$$

2. All raw material supplies from every source must be transported to the processing enterprises:

$$Q_r = \sum_i Z_{ri} \quad .$$

3. Raw material supplies from all sources must equal the quantity of raw material processed at all points of production of the finished product:

$$\sum_r Q_r = \sum_i x_i \quad .$$

The aim of the problem is to find the minimum total costs of transporting raw material to processing points, of processing it, and the amount of capital invested. Hence, the values of the unknown quantities must ensure the minimum of the functional:

$$\min\{\sum_i g_i(x_i)x_i + \sum_{r,i} a_{ri}Z_{ri} + E\sum_i U_i(x_i - x_i^0)(x_i - x_i^0)\}$$

When Stage Two has been completed, we will know the optimum variant, i.e. the variant that ensures:

(i) rational location of processing enterprises, with the best links (from the point of view of costs) with sources of raw material;

(ii) determination of the most appropriate capacities at each site, which enable all the available raw material supplies in each zone to be fully processed;

(iii) the minimum total costs of producing output and transporting raw materials.

Thus, solving the given sectoral problem in two stages covers the sphere of producing the finished product, transport of the raw materials to the processing enterprises, and transport of the product from one economic area or region to another. As a result of solving the problem, the optimum variant for the location of enterprises in the industry is known, i.e. the variant that ensures (a) minimum total expenditure on production of the finished product and on transporting raw material to the processing enterprises and delivering output to other regions, and (b) minimum expenditure on capital investment."

This quotation shows the role which this technique can play both within the planning structure in general, and in the specific determination of plant sizes.

The technique is also widely used by large corporations in the West, for optimizing current operations within the constraints of current equipment and sales potential. But it is less often employed in longer-term planning and capital investment studies (i.e. in many of the studies relating to scale). This is because of the environmental and competitive uncertainties already referred to. These oblige companies to give greater attention to adaptive and dynamic techniques, such as those reviewed next.

3.2.2 Dynamic programming (or a simplified version of it such as a "decision tree") has a number of features making it suitable in principle for the consideration of alternative decisions, particularly where those decisions will strongly constrain the future alternatives and the future resources available, and where there will be future decision points. This is typically the case where investment in a major unit of plant is considered. But it has a number of drawbacks:

1) The decision-criterion for "optimality" has to be one-dimensional--though the technique could be applied with several different criteria.

2) The "state-specification" may have to be multi-dimensional to represent with adequate variety the present or future situation of the organization; this may lead to major computational problems.

3) The statistical specification of the behavior of the future environment, and of its interaction with the firm, is a fundamental problem of this and any other technique.

4) A particular aspect of the environment is the behavior of other, independent organizations which may affect the future outcomes. This would require an extension into n-person "gaming" rather than the single-decision-maker technique.

3.2.3 Simulation as a technique is virtually unlimited in its breadth--since the word is almost synonymous with "model-building," but without any restrictive connotations of optimizing techniques. Multiple performance assessments could be made from any given "run." Interactive decision-making, and multiple decision-makers, could be incorporated. But the basic problems of modelling the environmental behavior cannot be readily overcome.

3.3 Developing "Standard Models"

The development of a "Standard Model" is characterized by the following steps:

(a) a survey of entities, similar in nature to that being contemplated, but varying in scale (note that this requires two key definitions: of "similar" and of "scale");

(b) a definition and evaluation of the performance characteristics of these entities;

(c) an attempt to develop a systematic relationship between "scale" and "performance" as defined in (a) and (b).

This we shall term the "standard model" approach, since its apparent aim is to develop a model, or "scale - performance relationship," which may be used as a standard of reference in considering future similar decisions.

Much depends on the use of the word "similar." The cautious exponent of this approach might add the words "mutatis mutandis" (having changed those things requiring to be changed). There are two major pitfalls. The first is the assumption of similarity between the group of entities studied in (a) above and the entity being considered now. By definition, the subject of current study is newer than the old sample, so there may have been technical change, even within superficially similar overall size characteristics. These could act either way on performance: either improving it, through better technology, or (at least initially) meeting new problems and difficulties because of learning problems and new complexities resulting from technological innovation. The new entity may in some dimensions lie outside the range of values in the original sample, and the extrapolation of the standard model's relationships may be technically invalid, even if it was correct for all values within the original data. Hu ttner [40] gives many examples of these faults in his review of the literature on electricity generation.

The risk of the "standard model"'s being incorrect when applied to a new situation is considerable, even where we are speaking of a purely technical model such as a heat balance equation in a chemical plant. It is even greater where the model is remote from basic physical laws, and is based purely on descriptive statistical techniques, such as regression analysis. We may characterize standard models on a spectrum "explanatory - descriptive," where a model is seen as more fundamentally "explanatory" - and, a priori, reliable and useful - the closer it is to the basic laws of science.

The second limitation of the "standard model" approach is that it is "environment-free," or tends to be so used. It is easiest to illustrate this by example. One might study, say, a number of farms, each characterized by its size, and construct various measures of performance. But any relationships emerging from such a study would have little value unless one either restricted the study (and therefore the applicability of its outcome) to a narrowly defined class of farms similar in climate and topography, soil type, degree of mechanization etc.; or built into the standard model explanatory factors relating performance not only to size, but to these other influences.

The second limitation may thus be overcome to some extent if the "standard model" can be developed to include explicit representation of relevant features of the environment (i.e. those affecting performance). Again, this is likely to be easier in the case of physical environment (e.g. rainfall statistics) than in the case of general economic and political aspects of the environment.

In general, "standard models" would seem safest in application at Level 1, the technical, "unit level."

3.4 Developing "Standard Methods"

The second way in which general conclusions are brought to bear on a specific decision is through the development of a standard methodology.

Since many methods centre on the employment of a standard model, the use of "standard models" may be seen as a sub-set of the broader class of "standard methods."

Every field of human knowledge develops methods and techniques of some generality, if only for the study and further development of that field of knowledge. In those disciplines which relate to the understanding and management of purposive systems, various techniques and methods have been developed, claiming some degree of general applicability to problems of management decision. Many of them are "uni-disciplinary," e.g. looking only at the financial dimension, or only at the implications for transport, etc. A systems analysis approach has some claim to the greatest degree of generality, although for many simple decisions, small in scale and localized in impact, such generality may be redundant.

One of the universally recognized features of problems of scale is that there is a need for change, and for increased sophistication of methodology, as one considers problems of larger scale: it is not simply a question of repeating the methods applicable at small scale, with all the numbers suitably multiplied.

Within a planned economy, small scale changes which do not alter the total volume of production are within the discretion of a factory manager, or the planning committee of a small republic; but major resource commitments must be considered within the overall plans.

Within a market economy, the building of two houses may be a matter for a small builder. He has some formalities with the local council for permission and compliance with standards, checks his position with his bank, buys the resources, builds the houses and (hopefully) sells them at a price covering his costs. The construction of 200,000 or 2 million houses requires more than a bigger builder, a bigger bank, and a bigger council. Nor is it merely the aggregation of many small builders, banks and councils. It becomes a multi-year, strategic question of national policy, resource availability, and the pattern of economic development and human settlement for decades ahead. The number of factors to be considered increases, as well as their individual magnitudes; and there must be corresponding development of methodology towards a "total system" approach, covering various dimensions, and various criteria.

One of the underlying sources of "problems" of scale is the tendency for decision-makers to apply methodologies, familiar and reasonably successful in small-scale decisions, to situations much larger in scale than those for which the methodologies were developed. This amounts to arguing by false analogy, often used in polemical situations. For example, in market economies, a large monopolistic organization may defend itself against government intervention or regulation by arguments in which the liberty of the individual small trader figures prominently, disregarding the qualitative differences to which the quantitative difference in scale and market power give rise. The distinctions of level drawn in 2.2 above should reduce the risk of such confusion, but more development of measurement is necessary: see 5.2 below.

The qualitative changes in environment have been described and generalized in a paper by Emery and Trist [26], which is described in 5.3 below. Some of the standard methods and models are reviewed below under "discipline" headings. But from what has already been said, the thrust of our "case for research" should begin to emerge in outline: it will lie in those areas where increases of scale are rendering existing "standard models" and "standard methods" less adequate, because the increase of scale at levels 1, 2 and 3 creates increasing effects at levels 2, 3 and 4. Changes at these higher levels, particularly changes of relevant context, may then change the assumptions used in the lower level calculations of scale.

3.5 The General Direction of the Required Methodological Developments

The preceding argument can be briefly summarized. Many mangement decisions, in government and other large-scale organizations, are concerned with the creation of larger entities than have previously existed in the contexts concerned. Such decisions typically differ from earlier decisions in ways which have implications for the methodology or models employed:

(a) the impact on the environment, and of the environment on the project, is greater; therefore

(b) more physical and economic variables have to be explicitly considered, and

(c) more individual and organizational parties have to be considered, consulted and bargained with; leading to

(d) the probability that there will be multiple, rather than single, criteria to be considered; and

(e) resulting from the planning and negotiating stages implied by the above, and from the complexity and scale of implementing the project itself, the time-scale to be considered will also have to be longer than previously.

Figures 7 to 9 convey something of the methodological development required. Figure 7 shows the basic elements: the decision-maker has resources, alternatives, and purposes: what can be called the "on" and the "for." There is an environment, which will constrain the alternatives and influence the evaluation; but at the level of problem where a "standard model" is applicable, the environment may be simply parametrized, e.g. by a single "growth rate of demand" figure, or "limit on available capital."

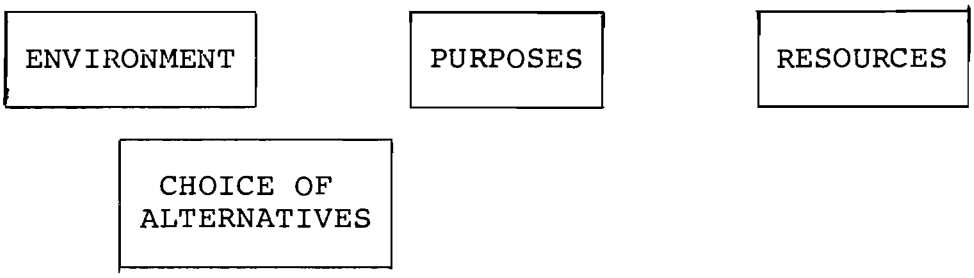

Figure 7: Basic Elements of the Problem

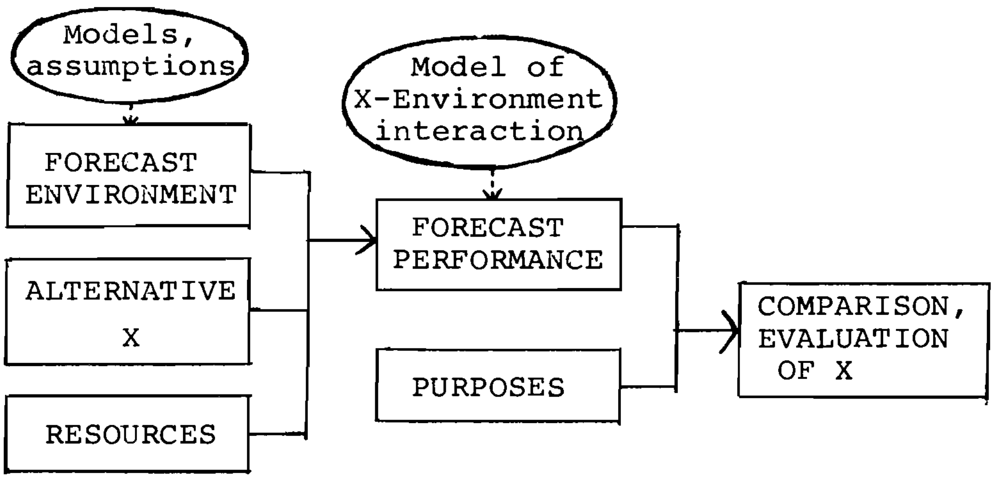

Figure 8: "Single-Shot" Evaluation Process

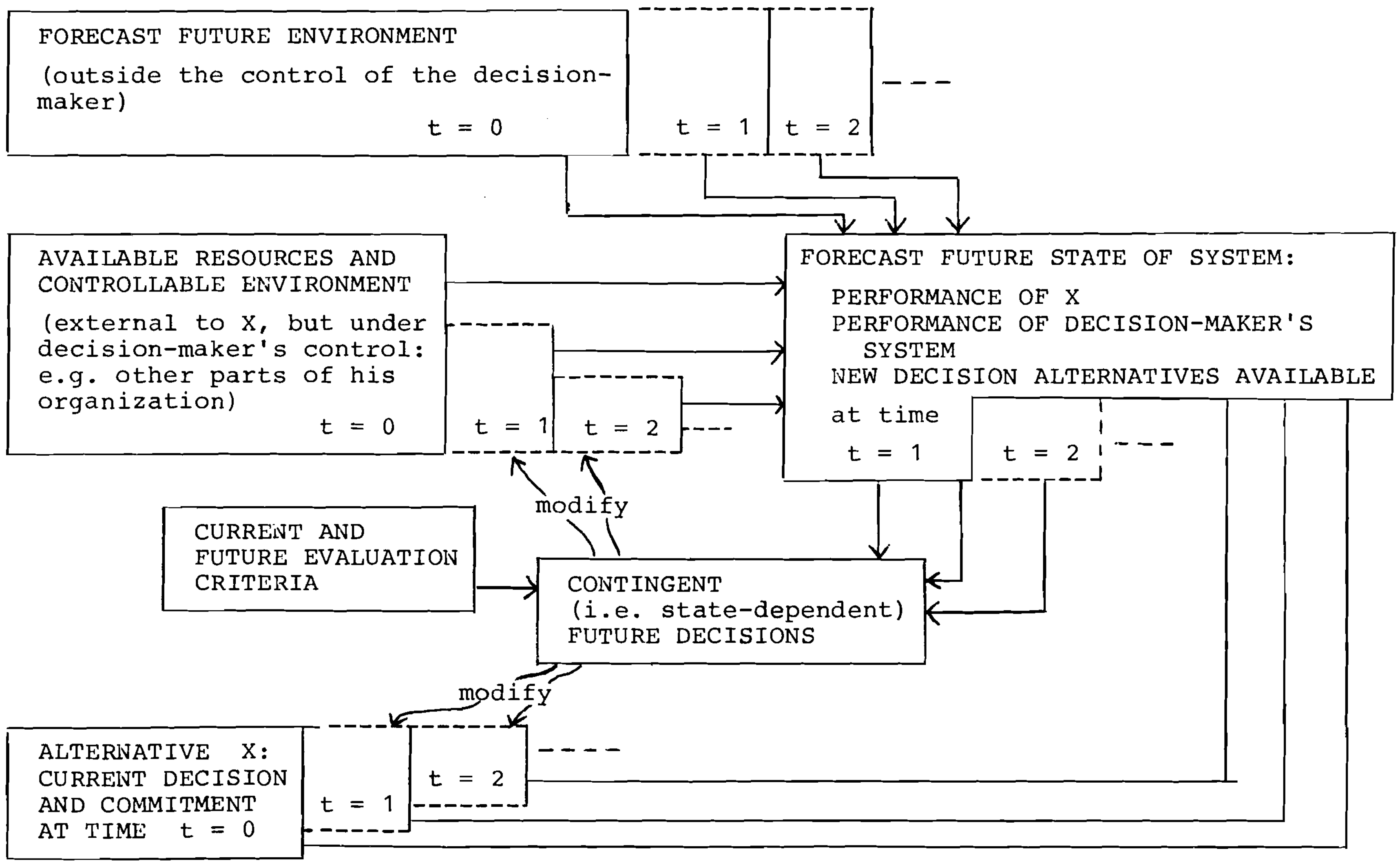

Figure 9: "Dynamic" Evaluation Process

In Figure 8, the evaluation process is made more explicitly dependent on the forecast future environment, and on the interaction between the chosen plan and the future environment. Such an approach might be appropriate where the decision and its implementation have to take place within a period of time short in relation to the time-span of its consequences, and without opportunities for reviewing the decision.

More typical of the situations relevant to increases of scale is the approach shown in Figure 9. This gives greater consideration to the dynamics of the situation, and the possibility of need for future modifications (since the time-scale has lengthened, there will be greater forecast uncertainty, and greater need, and opportunity, to review decisions). There is a greater recognition in Figure 9 of environmental effects: the big projects are significant features of their own future environments.

These figures and the preceding discussion provide a starting framework of reference and terms for a review of disciplinary approaches and techniques claiming relevance to "problems of scale."

4. DISCIPLINARY APPROACHES TO PROBLEMS OF SCALE

4.1 Introduction

Problems of scale in the real world do not arise under "disciplinary" labels like the papers in an examination. However, they have been studied from the perspectives of many different disciplines, and the approaches developed can in some cases be conveniently grouped under these headings. Some of the contributions of mathematics have already been introduced in the preceding section. The categories used below are somewhat arbitrarily divided, because subjects overlap, and the organization of subjects varies from country to country. Broadly speaking, within these categories, the authors reviewed in each relate and refer more to other authors within the category than to those outside it; they tend to use common terms, concepts and assumptions. The categories are labelled as follows:

- 4.2 Industry-Specific
- 4.3 Engineering and Technological Forecasting
- 4.4 Industrial Economics
- 4.5 Capital Investment Appraisal
- 4.6 Social Science Approaches to Questions of Organizational Scale
- 4.7 Human Settlements and Organization
- 4.8 Control Theory
- 4.9 General System Theory

4.2 "Industry Specific" Approaches

These embody the view that the problem of scale is so technical and industry specific that it must always be tackled entirely on an _ad hoc_ basis, in terms of a specific project, and that no useful generalizations can be made _from_ the project, nor do any general methods of analysis exist which can be brought _to_ the project; except from earlier, similar projects in the same industry.

The justification for such an approach is clear to the extent that an industry's product or technology is unique in some important respect: the non-storability of electricity; the perishability of newspapers or ice-cream; the characteristics tend to be commonest in primary industries, at the interface with natural resources. Thereafter, the industries of intermediate processing and conversion, and of transportation and distribution, tend to conform to general patterns; with product storage characteristics leading to some differentiation. Service and information-processing industries may require separate consideration, but their growth of scale and patterns of deployment show similarities to manufacturing industry in terms of organization, if not of technological content.

The "within-industry" literature tends to be oriented almost exclusively to the plant, (levels 1 or 2) rather than its environment, and to use simple or naive financial criteria uncritically. Bela Gold has documented these criticisms in many fields, and these are some of his conclusions at the end of a classic study of Japanese blast furnaces [36]:

> "... actual scale adjustments cannot be adequately evaluated within the limited perspectives provided by prevailing economic theory or common engineering approaches... More penetrating analyses of past or prospective decisions involving changes in scale would seem to require a broader exploration of the relationships between the array of expected benefits and burdens of scale adjustments and the array of basic managerial objectives."

This is not directly a plea for the development of common, cross-industry methodologies; but it does argue for the embodiment of the scale decision within the wider perspective of corporate strategic decision-making; a view already advocated in 3.5 above.

4.3 Engineering and Technological Forecasting

Engineering literature tends to be industry-specific, but there have been some cross-industry generalizations such as the concept of a "power-law": Cost = Capacity^k (e.g. k = 0.7), or the concept of the Reynolds number in hydrodynamics. On the former, Gold [36] comments sardonically:

"... the supposedly hard-headed engineering literature reflects long-standing and widespread acceptance of a 'rule' that each doubling of capacity tends to require increases in investment of only about six-tenths. Further inquiry reveals, however, that this expectation seems to be rooted solely in the simple-minded view that the volume increases more rapidly than the enclosing surface of rectangular, cylindrical and spherical shapes - and that the output of facilities tends to be correlated with their volume, while investment costs tend to be associated with the size of the enclosing surface. Such a relationship may hold, of course, in respect to some kinds of apparatus and facilities, especially in respect to the cost of constructing outer shells such as tanks, furnaces, boilers, pipes and simple buildings. But fundamental shortcomings narrowly restrict the range of its applicability."

Nonetheless, such superficial approaches continue to be widely propagated - e.g. Cameron [14] of Burmah Oil entitles his paper "Three Simple Steps to Determine Optimum Plant Capacity," and summarises his conclusions:

"(1) It is feasible to provide a basis for the optimum sizing of new plant using a routine which incorporates elementary economics.

(2) The principal quantifiable factors which determine the economic plant size are the anticipated market growth rate and the cost of capital.

(3) Large financial penalties can be incurred by undersizing or oversizing new plant.

(4) Optimum capacity solutions are characteristically robust and it is therefore possible to employ a broad brush approach in the treatment of plant capital and fixed costs.

(5) It is possible to derive an optimum plant investment cycle for specific financing situations."

Ball and Pearson [6] are technically more sophisticated in their approach to the engineering problems of scaling up size, but on the method of analysis of the investment decision quote Cameron [14] as "more than adequate" as compared with the "detail technique" described by the ICI authors [42] (discussed below).

An important aspect of problems of scale, to which the engineering studies pay greater attention, is that of their "multi-sectional" and "multi-functional" nature. One is not simply comparing black boxes of varying size. This has several implications of relevance to the methodology used in analysing such decisions. Technically, the scaling-up rules may be quite different for different parts or functions. In planning and construction, one might not necessarily build a perfectly "balanced" plant. Some parts might have a low marginal cost of extra capacity at the time of construction, but be impossibly expensive to expand in future years. Reliability considerations would point to redundancy, in number or size, of cheap but critical components. It might be possible to start the planning and implementation of a new plant's construction, but finalize its capacity specification a year or two later, in the light of latest information.

It is at the engineering level that such information has to be sought; unfortunately the published economic models and financial decision criteria tend to ignore these refinements.

The multi-functional view of a so-called "unit" of plant is of central importance to Gold [36], who suggests that "scale economies are derived from the increasing specialization of functions" and, hence, that "scale be defined as the level of planned production capacity which has determined the extent to which specialization has been applied in the subdivision of the component tasks and facilities of a unified operation." This is a strong and interesting proposal. According to Gold [36], it "raises doubts about the likelihood of finding scale effects which are universal among industrial processes covering the entire spectrum of physical and biological sciences, or over the entire size range of possible operating units within each." That Gold may be wrong here, at least in the context of biological units, seems to us clear: von Bertalanffy [8] provides many examples of "universal scale effects." But in industrial contexts, Gold's view is important, and he concludes "it would appear that major new horizons must be explored before new advances in our understanding of the generalizable and non-generalizable elements of changes in the scale of production are likely to be achieved."

While the views of those with detailed engineering expertise are rich in empirically-based understanding, it must be remembered that the technologist is not usually the best "generalizer" to use, for instance, in technological forecasting: his acquaintance with the "trees" can sometimes reduce his ability to view the "woods." There is, for instance, considerable evidence (reviewed by Sahal [65]) for the existence of "progress functions," "learning curves" or "experience curves," characteristic for each industry. These take such forms as "for every doubling in the cumulative total of items produced, there is a 20% reduction in unit cost." The authors of the SARU model [66] at the U.K. Department of the Environment take a similar view, corroborating research by Fisher [29]. Such "laws" are, if accurate, very relevant to the dynamic aspects of problems of scale, and have been so used by corporate strategists (see 5.2 below).

4.4 Industrial Economics

Industrial economists have long sought generalized models of input-output relationships in different industries, summarized by "production functions." Such investigations have often been technically deficient, for several reasons. For generality, a large sample is sought. This may lump together plants of different construction date, design, and other significant factors. The analysis is often at the level of the organization (level 3), rather than the unit of plant (levels 1 or 2), because of the greater availability of published information on (economic) performance. The mathematical models used are often over-simplified, e.g. Cobb-Douglas production functions (output = $x_1^{\alpha} x_2^{\beta} x_3^{\gamma}$) etc. where x_i are factors of production) because estimation of the coefficients can be done by standard procedures of linear regression, rather than because of any technical examination of the plant itself.

Gold comments:

> "Economic theory has long depicted 'scale effects' (i.e., the effects on minimum average unit costs of increases in the capacity of plants engaged in identical production activities) in the form of a U-shaped 'long-run' cost function. This represents the envelope of an array of U-shaped short-run cost functions which are assumed to show the cost output relationship of successively larger plants. Such elementary economic concepts have been widely diffused among engineers, businessmen and government officials and may well have encouraged receptivity towards proposals for continuing increases in scale. Unfortunately, however, analysis offers little support for the assumptions on which this theory rests. In addition to the usual assumptions of static economic theory - whose severely restricted purview is often overlooked in the course of enthusiastic efforts to make policy applications - the long run average cost curve rests on assumptions involving the universality of U-shaped short-term cost functions, the pattern of changes in their minimum cost points and the effects of changes in output level on the relative advantages of different-sized plants. Accordingly, continuing reliance on convenient assumptions in place of exploring the realities of industrial practice has rendered the traditional theoretical approach to scale economics widely inapplicable in concept and all but trivial in its posited effects."

Many of these criticisms are repeated, amplified, and supported by the evidence of other papers in Gold's 1975 book [37].

There is evidence in other contexts of the application of inappropriate, over-simplified economic models to large-scale decision problems involving aspects of scale and relevant context. Some of these are discussed in section 6 below.

4.5 Capital Investment Appraisal

Since most new units of plant (or major reorganizations) involve significant capital investment, scale alternatives may be appraised as alternative capital investment projects. A considerable literature exists on techniques for the appraisal of capital investments. Traditionally, industrial decision-makers are supposed to have used unscientific, subjective techniques or entrepreneurial "judgment"; or simple criteria such as "payback period": the length of time required to recoup the original investment. Academic criticism argued for the use of a more scientific and rigorous approach, based on the "cost of capital" or "time preference" (i.e. benefits now or later) of the decision-maker or the society. This appears to provide a clear-cut rule for comparing alternative patterns of future cash flows. The use of "discounted cash flow" (DCF) has therefore become widespread, and is, for instance, central to the published methodology of ICI Ltd. [42]. Constant discount factors have been used in socialist planning for many years.

But the adequacy of DCF has also been strongly criticized e.g. by Adelson [3], Meyer [54] and others:

> "When faced with a problem which extends over a significant time period, a time sufficiently long that we are no longer indifferent to the timing of cash flows and other events within it, we usually fall back on the simple technique of discounting to express our time preferences. We do so in spite of the complete absence of justification for discounting within the general framework of modern decision theory."
>
> (Meyer, quoted by Adelson)

They point out that while DCF is a rational and internally consistent technique for comparing a number of alternative cash flows, it says nothing about the process by which investment opportunities are created or identified, or about their inter-relationships between one another, or over time. This is an omission which in socialist planning is clearly overcome by relating capital investment to the sectoral, regional and national plans. But the criticism is valid in the context of market economies, and a similar critique is developed in the literature of corporate planning, for instance by Ansoff [5]. To treat alternative capital investments, such as plants of different scale, purely as financial transactions, is to ignore strategic implications of the alternatives which may be of far more significance to the long term objectives of the organization, including that of survival.

This criticism of the one-dimensional inadequancy of simple financial criteria is similar to the criticism Gold [36] makes of simple engineering or economic models of returns to scale, and brings one to the same conclusion: the need for a "broader exploration of relationships between...benefits...burdens...and

...basic objectives." In short, to seeing the scale decision in the context of total corporate strategy.

4.6 Social Science (Organization Theoretic, Managerial, etc.) Approaches to Questions of Organizational Scale

A considerable literature exists on the subject of organizations, their sizes, and various structural characteristics. This has only been briefly reviewed. Much of it is apparently descriptive, seeking general models and relationships (independent of the particular function or industry in which the organization participates). Such literature does not appear to be oriented towards application to specific decisions, although there is no reason why incisive descriptive studies should not be so used, if the descriptions include any measures of efficiency or effectiveness.

For example, one of the classic works is Alfred Chandler's [15] epic study of the growth of America's major corporations. This demonstrated the causal connections between certain types of industry and phases of their development (e.g. the railroads; Dupont Chemicals; General Motors), and the organizational forms adopted. Ansoff [5] drew extensively on Chandler [15] in his prescriptive work on corporate strategy.

Another researcher in the sociology of organization whose work advanced to a prescriptive stage was Joan Woodward [76]. Her work was particularly significant (and widely influential), because it appeared to display a systematic relationship between the technology of an industry and its optimum organizational form; with the implication that firms departing from this optimum would have poorer performance. This strong hypothesis has not been well supported by subsequent research, and Donaldson [21] claims that its results have been "disconfirmed." This critique was eagerly taken up and amplified by Eilon [25]. Attempts have been made to defend the original Woodward thesis at least at levels near the work flow: e.g. our level 2 rather than the level 3 of the total organization. At level 3, size appears to be the main determinant of an organization's structural characteristics; at levels 2 and 1, technology may be determining.

One of the most recent papers in this field (March 1978) is Dewar and Hage [20]: "Size, Technology, Complexity and Structural Differentiation: Toward a Theoretical Synthesis." Each of the four terms (size, etc.) is carefully defined in terms which are measurable, structural differentiation being considered both vertically (hierarchical levels) and horizontally (determinants). Technology is defined as "task scope." They are then measured for each of 16 social service organizations, in 1964, 1967 and 1970, thus giving data not only on the measures, but on their rates of change. Correlation and regression analysis are then applied, to try to determine associations and causal connections, and their relative strengths. For example, "Large organizations are and remain complex ones as are organizations with a variety of tasks. But are they both becoming large and adding more inputs at the same time? Which is the stronger causal process?" They found no effect of size on complexity, but suggest that perhaps "the amount of growth was not sufficient to generate the economies of scale necessary before additional administrative specialties could be hired." This type of interpretation is similar in concept to Gold's definition of scale as a function of degree of specialization (see 4.3 above). Size, rather than technology, is found to be the more important determinant of both vertical and horizontal differentiation. But the key paragraph of Dewar and Hage's paper acknowledges the extent to which such studies still fall short of operational value:

> "In considering the relationship between size, technology , complexity, and structural differentiation without considering the consequences of these relationships, this paper has dealt with only half of the story. One might well hypothesize that certain levels of differentiation, given a certain degree of complexity or kind of technology, would be appropriate or counterproductive in terms of other elements of social structure such as centralization and in terms of important organisational outputs such as efficiency or morale. It is unfortunate that Woodward's (1965) lead has not been pursued in much of the recent literature. In what is perhaps the most interesting figure in her book she points out that there is apparently an appropriate span of control for a given technology, if performance is to be maximized (Woodward, 1965: 70-71). The implications are that by adding into future studies sets of performance measures, the field of complex organizations may be able to substantiate the Lawrence and Lorsch (1967) insight that unless there is a balance between differentiation and integration, productivity and effectiveness may suffer; that if one desires a certain set of outcomes, there are appropriate degrees of vertical and horizontal differentiation given the existing technological constraints; and that certain structural arrangements facilitate certain kinds of control and coordination while others hinder them."

The literature to which the above is a brief introduction is obviously important to any general study of problems of scale in organizations. It is empirical, quantitative, and seeks generality. But it may be problematical to translate conclusions from public service organizations to situations in manufacturing industry; and the dearth of studies including comparative performance measures is a serious deficiency.

To incorporate into these models questions of scale at the level of the technological unit, one would need a means of translating scale alternatives into their alternative organizational implications; it is not yet evident that any rigorous way has been found of doing this, or even whether any such unique relationships need exist.

4.7 Human Settlements and Organization

In a general discussion of problems of scale, some mention should be made of patterns of human settlement. Much of the history of civilization is related to overcoming the problems of coping with successively larger communities, both in terms of local settlements and at national and supra-national level. Single areas of settlement now range over six orders of magnitude in their population: from isolated houses to cities of several million people.

The diversity of circumstances, and the obvious fact that these are specific satisfactions and drawbacks associated with every size, show at once that there is little point in seeking any simple solution to "optimum size" or "optimum mix." There are descriptive models of urban growth, and this literature has not yet been extensively surveyed; Forrester's [30] attempt to apply a "System Dynamics" simulation model to the city was not a very successful example of attempting to carry over simple analytical models into urban planning. The professional urban planners do not themselves appear to have developed clear views on either desirable target patterns, or standard and satisfactory methods, for land-use and urban planning. Coleman [16] documents the apparent failure of British post-war land-use planning. Some of the regional strategic plans in the U.K. have also drawn heavy criticism of their unimaginative and over-simplified techniques. A general review of the current state of Urban Planning theory, problems and models is provided by Winger [75]. In a critical and pessimistic article, Schneider [67] remarks that "planners operate without a conception of an ideal city. Especially in the United States, there is no established norm for size, either with upper or lower limits. There is no economic ideal, no formula for urban productive or consumptive efficiency."

A feature dominating any normative or prescriptive approach to planning the scale or pattern of human settlement is the extent to which it is dominated by the existing pattern. The rate of significant possible change is normally so slow, that major change can be achieved only over many decades. To forecast and plan for many decades ahead demands heroic assumptions about the uncertainties, or reflects a scarcely justifiable attempt to create certainties to which other future events must themselves adjust.

The scale of towns and cities is principally of significance to other decisions about scale, in that the former often define the environment within which the latter are made. This is equally true of national environments, for activities to which this is the relevant measure; and there have been many scale-related arguments for the creation of international activities and supra-national entities: to these we return in 6 below.

4.8 Control Theory

The literature of control theory contains many contributions from electrical and electronic engineers, and from mathematicians and cyberneticians. It concentrates on technical situations, amenable to analytical modelling, computer simulation, and technical experimentation. Its application appears to have been very local in origin - the control of automatic machinery or process plant. But increasingly there have been attempts to extend the scope of the formally structured control systems to larger systems, such as an integrated steelworks complex; and at least on a theoretical level, the methodology has been applied to larger scale problems such as economic management.

The subject has not been extensively reviewed within the current project, but is here noted for the sake of completeness, and with an awareness that it has much to contribute on the methodology of formal control in certain types, and on certain scales, of organization. A useful starting point is the April 1978 Special Issue of the I.E.E.E. journal, "Transactions on Automatic Control." This issue is devoted to "Large-Scale Systems and Decentralized Control," and in his editorial reviewing the content of the issue, Athans makes the following significant general observations:

> "... We are observing the formation of several schools of thought in regard to large-scale systems, and I believe that these schools of thought are well represented by the papers in this issue.
>
> It should be self evident that the coordinated control of complex man-made systems will represent the great challenge for the next several decades. In a world of limited and dwindling resources, we can see a greater need for optimization, often under conflicting and fuzzy performance criteria. At the same time we see a greater interconnection between systems. The global economic system is an example of this, in which the economic policies of one nation can have significant impact upon the economic welfare of several other nations. If we turn our attention to physical systems, we can see several examples in which existing large-scale systems operate in a relatively inefficient way due to poor planning, lack of systematic decentralized yet coordinated control, and failure in emergency situations. In the area of power systems, we see an increasing degree of interconnection, with subsequent ill-understood dynamic phenomena, which can result in severe blackouts. Large-scale transportation networks are a mess; consider the dubious effects of diamond lanes, and the failure of deterministic scheduling algorithms to function effectively in a dynamic stochastic environment encountered in recent "dial-a-ride" demonstrations. In the area of complex data communication networks, such as the ARPANET, only

about 30 percent of the network resources are used to transmit real information, while the remaining 70 percent are used to transmit protocol (control) information. Sudden changes in demand and failures can set up dynamic instabilities. In the area of batch manufacturing, involving metal cutting by several interconnected machines, recent U.S. statistics show that the machines cut metal about 3 percent of the time, while over 90 percent of the time the metal parts are either moving from machine to machine or gathering dust in queues. ...

The inefficient operation of large-scale interconnected physical systems can be attributed to lack of fundamental understanding and modeling of the underlying interactions, the lack of coordinated control strategies, and the use of deterministic static strategies in an inherently dynamic and stochastic system. In view of their basic training, systems engineers and scientists have a lot to contribute toward improving the efficiency, productivity, and reliability for such complex systems.

If systems theorists are going to have a significant impact toward improvement in the operation of such complex systems, then they must, by necessity, become more interdisciplinary in their outlook. Closer interaction with operations researchers is necessary, since many of the complex systems have an inherent network structure; existing results in complex multicommodity flow network problems will have to be extended to the stochastic case; and dynamic interaction phenomena will have to be understood. The need for and cost of communication channels, their fidelity, and the impact of delayed information on decentralized decision making is also an essential part of the problem. One needs to make precise the value of information for real-time control. In this respect, interactions with communications engineers and information theorists is important in an attempt to, perhaps, extend the noncausal aspects of information theory to the causal requirements of real-time control. With respect to information, one must take into account the distributed sensors, the need for decentralized estimation, the storage of information in distributed data bases, and decision making using distributed computation. For these reasons, increased interactions with computer scientists is extremely important. Finally, one must not forget that reliable operation, in the presence of several and possibly simultaneous failures and/or abrupt changes in the underlying system, is crucial. A theory that allows us to compare classes of decentralized information and decision structures, and eliminate inferior ones on the basis of reliability, would be extremely useful.

The difficulty of developing the new theoretical tools for decentralized control for large-scale systems should be recognized. Even for centralized multivariable problems, we are only now beginning to stand their properties in terms of robustness, integrity, failure management, and reconfiguration. Decentralized multivariable control problems promise to have a multifold complexity. Using traditional optimality considerations, one is faced with great complexity because of nonclassical information patterns. The recent experience in stochastic dynamic teams and games shows the great complexity of stochastic control strategies associated with different solution concepts, e.g., minimax, Nash, Stackelberg, etc. On the other hand, if one models appropriately physical phenomena commonly encountered in large-scale systems, e.g., time-scale separation, weak coupling, etc., then perturbation methods coupled with existing theory can result in decentralized structures. This points out that careful interplay between physical problems and theory is necessary for the development of relevant theory and algorithms.

It is my opinion that, from a theoretical point of view, we have almost exhausted the power of existing methodologies and theories. It should be noted that traditional servomechanism theory as well as the tools of modern control theory (such as the maximum principle, Lyapunov stability theory, estimation theory, and dynamic programming) represent centralized design methodologies. These can be extended to a certain degree to attack important problems for large-scale systems, as can be evidenced by the contributions to this Special Issue. What we need from a theoretical point of view are novel and innovative approaches for comparing alternate decentralized information and decision structures. The current state of the theory does not allow us to do this. The new theories will have to bring in new concepts of solutions, new definitions of what we mean by optimality, with special emphasis on reliable operations, and a more fundamental understanding of the value of information for decision making. In short, we need brand new theories for the future, and this is why the field of large-scale system theory and decentralized control will continue to be an exciting area for both theoretical and applied research in the decades to come."

4.9 General System Theory

Although it would seem to be the natural background or basic philosophy of systems analysis, general system theory does not appear to have won the widespread acceptance or familiarity to which its claims of universality might have entitled it. The term is closely associated with the name and work of von Bertalanffy, and the Society for General Systems Research follows in this tradition, publishing a journal with the somewhat misleading title "Behavioral Science," and an annual yearbook (edited for many years by Rapoport).

Von Bertalanffy is pre-eminent in demonstrating or asserting the underlying similarities of structure between superficially dissimilar systems and concepts, and is confident of the suitability of the general system theoretic approach to the study of organizations. He cites the work of Boulding [11]:

> "As an example of the application of general system theory to human society, we may quote a recent book by Boulding, entitled *The Organizational Revolution*. Boulding starts with a general model of organization and states what he calls Iron Laws which hold good for any organization. Such Iron Laws are, for example, ... the law of optimum size of organizations: the larger an organization grows, the longer is the way of communication and this, depending on the nature of the organization, acts as a limiting factor and does not allow an organization to grow beyond a certain critical size. According to the law of instability, many organizations are not in a stable equilibrium but show cyclic fluctuations which result from the interaction of subsystems. ... The important law of oligopoly states that, if there are competing organizations, the instability of their relations and hence the danger of friction and conflicts increases with the decrease of the number of those organizations. Thus, so long as they are relatively small and numerous, they muddle through in some way of coexistence. But if only a few or a competing pair are left, as is the case with the colossal political blocks of the present day, conflicts become devastating to the point of mutual destruction. The number of such general theorems for organization can easily be enlarged. They are well capable of being developed in a mathematical way, as was actually done for certain aspects."

This type of view of the general behaviour of organizations has similarities to Emery and Trist's work, described in 5.3 below. Von Bertalanffy himself started work as a biologist, and continues to use many of its laws as being of wider applicability:

> "Relative Growth
>
> A principle which is also of great simplicity and generality concerns the relative growth of components within a system. The simple relationship of allometric increase applies to many growth phenomena in biology (morphology, biochemistry, physiology, evolution).

A similar relationship obtains in social phenomena. Social differentiation and division of labour in primitive societies as well as the process of urbanization (i.e. growth of cities in comparison to rural population) follow the allometric equation. Application of the latter offers a quantitative measure of social organization and development, apt to replace the usual, intuitive judgments (Naroll and Bertalanffy, 1959). The same principle apparently applies to the growth of staff compared to total number of employees in manufacturing companies (Haire, 1959)."

From the point of view of a systematic study of problems of scale, it is doubtful whether any ready-made answers can be lifted directly from biology - e.g. a facile translation of "the reasons for the extinction of dinosaurs" into "the problems of large technological units." But the prospect remains an intriguing one, and the possibility of developing better methods and perspectives in systems analysis out of theoretical and conceptual developments in systems theory remains open, and strongly argued (e.g. Weinberg [74], Ackoff and Emery [2]).

5. RESEARCH: METHODS, ISSUES, MATERIAL

5.1 Introduction

The logic of this paper has been towards approaching problems of scale through a mixture of socialist economic planning, corporate strategic planning and system theory, drawing freely from any disciplines offering useful insights, and seeking to learn from their shortcomings.

To move towards a practical research programme, we consider next (5.2) the very basic question of measurement. In 5.3, a typology of environments is introduced, as a basis for defining classes of scale problems. In 5.4, a general discussion of long-term dynamics of scale problems includes examples of a wide range of situations in which scale problems occur.

5.2 Measurement of Size

This must at present be viewed as an "area for further research." If one is seeking to identify the stage of growth at which the need arises for a change of techniques and methods, and to identify this stage in different industrial and social contexts, then one wants, ideally, measures of scale which are independent of the specific area, and comparable between areas.

One possibility is to focus on absolute quantities which are meaningful across many areas, and not specific to one industry: e.g.

Number of people employed

Physical area or volume occupied

Physical mass or volume of annual throughput

Financial value of the capital employed

Financial value of annual output

Another possibility is to use relative quantities, such as the following types of ratio:

(a) $\frac{\text{size of unit being considered}}{\text{size of largest existing unit}}$

or

(b) $\frac{\text{size of unit being considered (capacity, annual output)}}{\text{size of relevant context.}}$

The second of these raises a fundamental problem of definition - what is the relevant "context" for deciding whether a unit is relatively "large"? It could be a world total, a national or regional total, or a total within the one organization. Like the word "strategic," the term is relative. The relevance of different base scales depends upon the degree of interaction between region/country/world, etc.: a low value per ton product (e.g. quarried stone) would usually have a more local context than a high value per ton product (e.g. semi-conductors). Relative or absolute decline in transmission, transport and/or communication costs may change the relevant boundaries, as can political decisions on the control or de-control of trade flows.

Simmonds [70 & 71] has published papers containing carefully researched, empirical studies of scale effects in the Canadian and U.S. chemical industries. In these, he uses as a key measurement the ratio of the "largest single-train plant" (i.e. the largest which depends on one major component) to the total market or production of a country (whichever is the larger.) His evidence is that "the size of the largest plant has usually kept pace with the growth of the market." In his second paper, he uses an examination of relative scale and scale economies to consider the comparative competitive position of Canadian and U.S. firms in the Canadian market; and shows that "across-the-board percentage tariff reductions are ineffectual for industrial nations with relatively small domestic markets such as Canada, in major products such as petrochemical intermediates." Simmonds also points out the various scales of definition of "market," which indicates some of the problems of measurement and specification arising in the definition of "relevant environment."

In the same context, it is of interest to note the use (without definition) of the term "world-scale plants": "The cornerstone of our investment planning is to establish world-scale plants." This quotation is from The Chairman's Report, 1977 of the large U.K.-based chemical corporation, Imperial Chemical Industries Ltd. A feature of scale effects is a general enlargement in the geographical scale of "relevant contexts." A number of further examples, raising the question of "efficiency v. self-sufficiency," are discussed in section 6.

A related topic is the use in a competitive context of "relative market share" (i.e. a measure of type (a) above) as a measure of an organization's strategic strength. This, allied to concepts of "product life cycle" and "experience (learning) curve," has been extensively propounded by the Boston Consulting Group as a basis for strategy formulation. Delombre and Bruzelius describe a case study from SKF group, a multi-national company operating in the field of precision engineering. Their conclusion is that "the correct measure of competitive posture...is own market share/market share of biggest competitor." The logic is that the greater experience leads to lower costs; and cost "is relative, not absolute...no one knows what a cost ought to be ...the low cost can at any point in time only be defined by the company which has achieved the lowest cost so far." Here scale is being measured by cumulative production (= experience) rather than by unit capacity or size of firm.

This is not a digression from the subject of this sub-section, "Measurement." The point is that the types of measure relevant to the study of problems of scale will often be relative measures; that is to say, the measures will be properties arising not only from the entity under consideration, but from its relationship to its environment. Thus even if the organization stands still, changes in its environment may change its scale. We consider next the question of generalized description of environment.

5.3 The Changing Environments

In defining the problems of scale in general terms, this paper has sought to emphasize two particular aspects inadequately treated in much (though not all) of the existing literature: the consideration of the environment; and the consideration of the dynamic behavior of the combined system of the entity under study and its environment.

To pursue this line of thought requires the development, on a general level, of conceptual models of the nature of the environments within which problems of scale are typically considered. In addition, we shall be interested in considering at a general level those changes in the nature of the environment in recent years which have stimulated or necessitated changes of scale in operating units or organizations. The fundamental differences between the environments in planned economies and market economies have been referred to in 3.1; but relaxation of international tension, increasing East-West trade and long-term agreements mean that each system needs to develop greater understanding of some of the characteristics of the other.

Chairman of the USSR Council of Ministers, A.N. Kosygin described this phenomenon in the following way [47]:

> "In the conditions of detente new qualitative aspects are being acquired by our economic relations with the developed capitalist countries, relations that can develop successfully on the basis of the principles set forth in the Final Act of the Conference on Security and Co-operation in Europe. We shall continue the practice of signing large-scale agreements on co-operation in the building of industrial projects in our country and on the participation of Soviet organisations in the building of industrial enterprises in Western countries. Compensation agreements, especially those covering projects with a short recoupment period, various forms of industrial copoperation and joint research and development are promising forms of co-operation.
>
> Of course, our trade and economic relations will develop faster with those countries which will show a sincere desire for co-operation and concern to ensure normal and equitable conditions for its development. Only in this case is it possible to maintain really broad and durable economic relations, which will be reflected in our economic plans."

The environments in planned economies were described in 3.1. we consider now an interesting attempt to give a general description of environmental changes in the Western, market environments.

5.4 The Emery and Trist Environmental Types

An important attempt to create a general "typology of environments" was that by Emery and Trist [26], and because of its potential relevance we reproduce here the concluding section of their paper, summarising four different environmental "types."

Summary of Emery and Trist's paper, "The Causal Texture of Organizational Environments":

"1. A main problem in the study of organizational change is that the environmental contexts in which organizations exist are themselves changing - at an increasing rate, under the impact of technological change. This means that they demand consideration for their own sake. Towards this end a redefinition is offered, at a social level of analysis, of the causal texture of the environment, a concept introduced in 1935 by Tolman and Brunswik.

2. This requires an extension of systems theory. The first steps in systems theory were taken in

connection with the analysis of internal processes in organisms, or organizations, which involved relating parts to the whole. Most of these problems could be dealt with through closed-system models, such as that introduced by von Bertalanffy, involving a general transport equation. Though this enables exchange processes between the organism, or organization, and elements in its environment to be dealt with, it does not deal with those processes in the environment itself which are the determining conditions of the exchanges. To analyse these an additional concept - the causal texture of the environment - is needed.

3. The laws connecting parts of the environment to each other are often incommensurate with those connecting parts of the organization to each other, or even those which govern exchanges. Case history I illustrates this and shows the dangers and difficulties that arise when there is a rapid and gross increase in the area of relevant uncertainty, a characteristic feature of many contemporary environments.

4. Organizational environments differ in their causal texture, both as regards degree of uncertainty and in many other important respects. A typology is suggested which identifies four 'ideal types,' approximations to which exist simultaneously in the 'real world' of most organizations, though the weighting varies enormously:

   a. In the simplest type, goals and noxiants are relatively unchanging in themselves and randomly distributed. This may be called the placid, randomized environment. A critical property from the organization's viewpoint is that there is no difference between tactics and strategy, and organizations can exist adaptively as single, and indeed quite small, units.

   b. The next type is also static, but goals and noxiants are not randomly distributed; they hang together in certain ways. This may be called the placid, clustered environment. Now the need arises for strategy as distinct from tactics. Under these conditions organizations grow in size, becoming multiple and tending towards centralized control and coordination.

   c. The third type is dynamic rather than static. We call it the disturbed-reactive environment.

It consists of a clustered environment in which there is more than one system of the same kind, i.e. the objects of one organization are the same as, or relevant to, others like it. Such competitors seek to improve their own chances by hindering each other, each knowing the others are playing the same game. Between strategy and tactics there emerges an intermediate type of organizational response - what military theorists refer to as operations. Control becomes more decentralized to allow these to be conducted. On the other hand, stability may require a a certain coming-to-terms between competitors.

d. The fourth type is dynamic in a second respect, the dynamic properties arising not simply from the interaction of identifiable component systems but from the field itself (the 'ground'). We call these environments <u>turbulent fields</u>. The turbulence results from the complexity and multiple character of the causal interconnections. Individual organizations, however large, cannot adapt successfully simply through their direct interactions. An examination is made of the enhanced importance of values, regarded as a basic response to persisting areas of relevant uncertainty, as providing a control mechanism, when commonly held by all members in a field. This raises the question of organizational forms based on the characteristics of a matrix.

5. Case history II is presented to illustrate problems of the transition from type 3 to type 4. The perspective of the four environmental types is used to clarify the role of Theory X and Theory Y as representing a trend in value change. The establishment of a new set of values is a slow social process requiring something like a generation - unless new means can be developed."

(Case history I concerned a company in the U.K. food-canning industry; Case history II concerned a total industry and its relations with society: the National Farmers' Union of Great Britain.)

The relevance of the above analysis to the consideration of scale is shown at several points. In type 'a', "organizations can exist adaptively as single, quite small units." This corresponds historically to a primitive stage of economic organization. In type 'b,' "organizations grow in size," because their size enables them to exploit environmental features more effectively. Thus the wider scale "all Soviet Union" basis gives greater total welfare and strength to each republic; the financial scale of a large firm gives it access to areas of high expected return, where the risks would preclude small organizations (e.g. in banking, insurance, or in areas where capital intensity and growth in scale give access to low operating costs). In type 'c', the dominating feature is competition, and this type appears applicable to developed market economies. Scale becomes perceived as an instrument of competitive strength, as in military contexts (e.g. Lanchester's laws on a tactical level) or in terms of market share, as in the strategy analysis based on "dominant market share" referred to in 5.2.

In level 'd', the "turbulent fields," the interactions and combined activities of the organizations, however large, contribute to effects beyong their control or expectations, thus altering their environment. It is not difficult to see the examples of this in industries such as steel, fibres, or ethylene, where the combined investment decisions of the major producers produce disastrous commercial results. (Simmonds [71] demonstrates the effect in chemicals). Similar effects occur in industries such as whaling, where the target catches of a few large and determined participants may exceed the sustainable yield of the field, with consequences ultimately disastrous for all. Parallels could be drawn in many other areas of natural resource exploitation or expropriation (e.g. land enclosures in Britain, 1780-1820; the current conferences and debates on maritime territorial rights; political conflicts in Africa; the strategic arms race; the cumulative effects of competitive consumer advertising on the minds of a "television-intensive" population).

It is characteristic of many of the examples quoted that the response has often been the attempts by the participants to establish and reinforce commonly held values as a constraint on their behavior. Examples are fishing quotas, Strategic Arms Limitation Talks, OPEC pricing, or the development of cartels in oligopolistic industries. This is as predicted by the Emery and Trist reference to "common value systems." In many industries, the development of large scale organizations and/or the deployment of large scale technological units may, whether or not the apparent decision-makers consciously intend it, represent transitional steps towards a stage of development in which common acceptance of increased constraint and regulation, and joint planning of future activities, will become unavoidable.

This type of general analysis and discussion tends to appear over-philosophical and speculative, as soon as it leaves concrete operational realities. But in our opinion it may provide a route to greater general understanding, and to the development of more appropriate methodologies, in many of the contexts in which problems of large scale organization appear. Some of these contexts are illustrated in the following sections.

5.5 The Need for a Research Framework

The previous sections have outlined many different approaches to problems of scale, and have introduced such terms as levels; factors; the measurement of scale; the description of different environments. But as a framework for research, we should seek to develop some conceptual picture of how these various terms are related to one another, and how these relationships change with time or depend on identifiable factors.

As one example, an outline picture is shown in Figure 10 of the way in which one might model the historical evolution of certain industries in the Western market economies.

This shows the "mechanism," or system, or process, by which the scale of units, plants, organizations and their relevant contexts have increased.

The double-lined boxes are the partially unknown, unbounded "environment" of the whole system comprising:

(a) new technical possibilities, as yet unknown, undeveloped or unimplemented;

(b) the potential demand for the final output of the system.

The rounded boxes represent "behavioural" elements of the system; the rectangular boxes, the identifiable and measurable effects. The distinction is in some cases not clear-cut.

There are in this diagram some areas of the system well understood and measured; others speculative, uncertain, requiring further research. The diagram attempts to put together an outline picture of the "total system," though for simplicity it omits competitive interaction, which in type 3 and type 4 of the Emery and Trist environments is a crucial stimulus to growth.

For example, the "two-thirds power law" would be one element linking increase of scale to cost reduction. Within the current "relevant context," this could lead to rationalization into fewer, larger units. It could also enlarge the relevant context, not only of individual units, but of the whole organization or industry, since the improved performance may increase the acceptability to a wider area. The general increase in scale of total activity in turn has a number of effects. The increase in cumulative production may, especially in a new and rapidly growing area, lead to improvements in technique and cost reduction; this is the theory of the "Technical Progress function" (see 4.3 above), though it could be cause and/or effect of unit scale increase. The view documented by Simmonds (and no doubt others) that a constant ratio holds between total market and maximum unit size may be taken as an additional or alternative hypothesis. The direction of causality seems more likely to be as indicated.

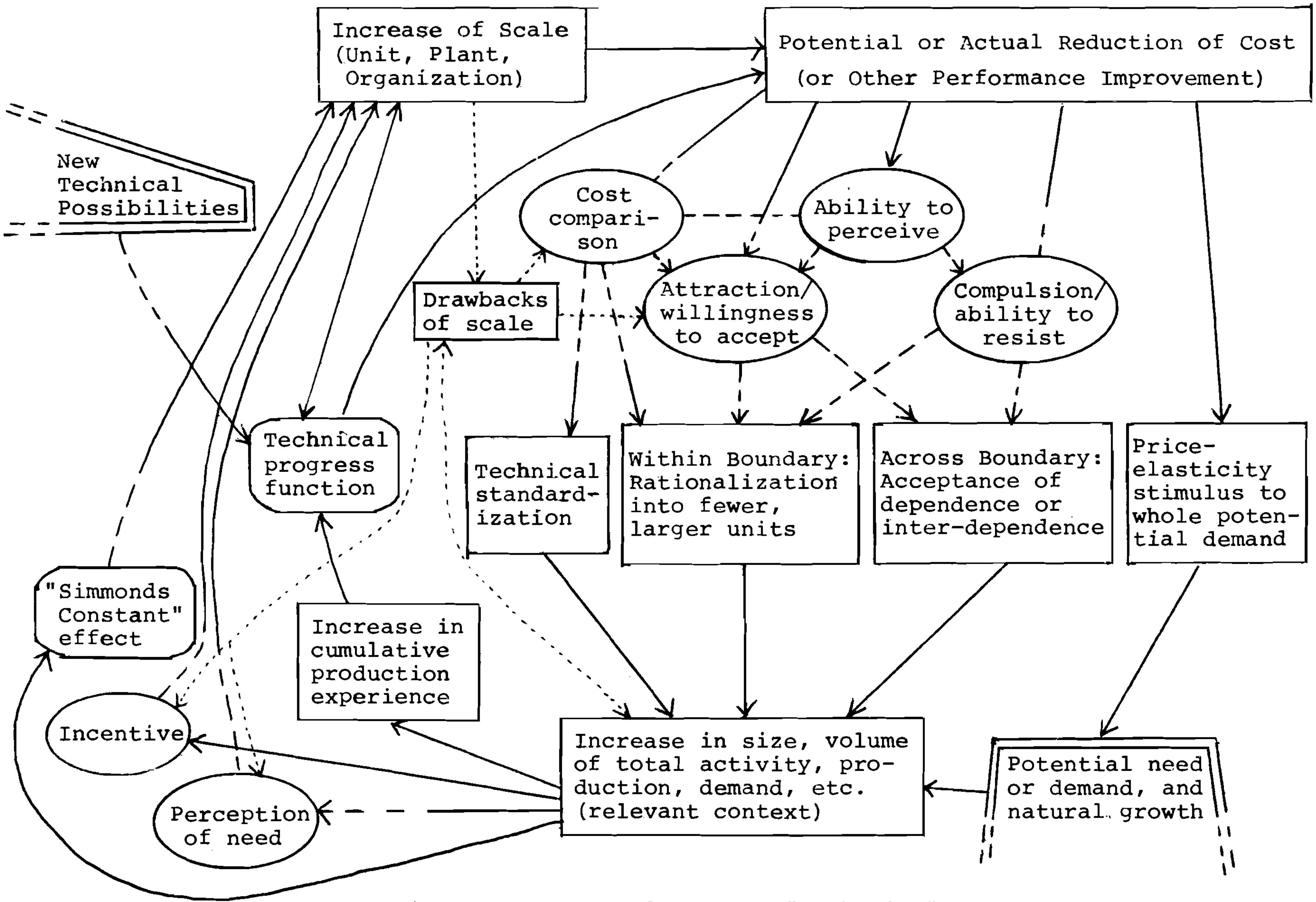

Figure 10. The Scale Growth "Mechanism"

Figure 10 could be further elaborated in theoretical detail. The relationships and assumptions could be examined in more detail, tested in specific contexts. "Drawbacks of scale" have not been included in any detail. In short, it offers a preliminary framework for research.

The evolution of larger scale plants in the socialist countries has taken place in a different environment, and under different objectives. But there would be some similarities at a technical level. In defining the framework for research at IIASA, one would seek particularly those aspects of the total scale and environment system which are common to many industries, and to East and West.

5.6 Research Material

Section 1 reiterated a basic principle of applied systems analysis as being that methodological development should arise out of the study of real, current problems; rather than being pursued as an academic activity. This is reflected in Table 1, the structure of the area's research programme. The following three "sources" of problems, or bases for seeking and selecting problems, can be considered.

(a) Problems already offered by contacts through National Member Organizations.

(b) Problems chosen to correspond as closely as possible to the areas of need identified or indicated in the preceding sections.

(c) Problems chosen to match the interests, experience and capabilities of area staff.

Ideally, all these three will coincide. In practice, they are bound to differ to a greater or lesser extent. In section 6, examples are given of case problems selected or proposed in each of the above three categories.

6. CASE STUDIES OF RELEVANCE TO PROBLEMS OF SCALE

6.1 Introduction

The following case studies, or potential case study areas, are suggested for their relevance to the study of problems of scale. This does not imply, however, that scale is the most important or central question in these case studies: the methodology and central concerns would always have to be subject to the needs of the situation as it was increasingly understood in the course of the study.

6.2 Electricity Generation: the Scale of Plant

The co-operation of one IIASA member country has been offered in a study of the question of the scale of electricity generating units to be installed at their next major coal-fired power station. Background information and discussions about the project are described in a series of internal working notes, and these details will not be repeated here.

This case study has a number of advantages - particularly the fact that contacts have been established, preparatory investigations made, and a start made on collecting and studying the relevant literature. Also in favour of this industry as an object for study is the apparent existence of substantial economies of scale, but also of growing doubts about how far these are in practice achievable beyond a certain point. A short review of some of the relevant literature will indicate some of the areas of controversy, which are closely relevant to current decisions.

There is a well-documented historical evolution of ever larger coal-fired generating units (to over 1000 MW). Landon [48], for instance, states:

> "The existence of substantial scale economies in fossil steam generation, up to the largest sizes with which we have a statistically valid experience, cannot be denied."

Landon quotes in support of this view the work of his colleague, Huettner [39]. Huettner has undertaken careful and critical review of the work of several economic studies of electricity generating units, analysing their deficiencies, and while he supports the case for scale, does so with some caution.

> "Since 1930, all of the long-run average cost curves have been L-shaped. More important, all indicate that economies of scale decline very rapidly but do persist throughout the observed range of plants sizes. In fact, from 1951 to 1968 the unit capacity costs and unit operating costs of 300 MW plants were never 10 percent higher than those of the largest plants constructed at the same point in time. The failure of previous studies to recognize the sharp reduction in scale economies for generating plant sizes above 300 MW may be due, in part, to their failure to include a sufficient number of 400 MW and larger plants in their samples."

The case for scale was long accepted in the U.K. electricity generating industry, e.g. as documented by Brown and Booth [13] or Booth and Dore [10] over 20 years ago. Again in a recent paper, Lee [49] confirms:

"My analysis clearly shows that there is economy of scale in both components and plant construction. By taking advantage of this through centralization, the industry has been able to continually reduce the price of electricity in the past 20 years."

Lee acknowledges, however that:

"In spite of our limited knowledge, there is enough evidence to indicate that there is an optimum size because:

Economy of scale is not constant; it decreases with increasing size.

There are some size-related causes of unavailability."

Landon argues that the U.S. electricity generation industry has been slow in adapting the larger plants, and in adopting the larger-scale organization required to achieve these economies. He therefore argues for a change in government regulation, to stop safeguarding by anti-trust legislation the smaller utility companies, and to facilitate rationalization. This argument appears to jump from the level of the generating unit to the level of the national system, without considering the effect of organization. Nerlove's analysis of utility companies fills this gap, concluding [56]:

"... that there is evidence of increasing returns to scale at the firm level in U.S. steam-electricity generation, but that the degree of returns to scale varies inversely with output and is considerably less, especially for large firms, than that previously estimated for individual plants."

Huettner [39] acknowledges that his and many other studies have concentrated on generating costs alone, whereas "planning and decision making for plant sizes, plant locations and transmission facilities are done at the system level." He refers to the difficulty of conducting system studies ("usually simulation analysis is required"). The methodology for such system simulation is, however, now well-established in most electric utilities and the development of computing facilities has probably reduced the cost. The essentials of a system simulation were comprehensively described as long ago as 1958, by Schroeder and Wilson [68].

However, the system simulation studies referred to are essentially "operational" simulations, evaluating the performance, under various load conditions and plant assumptions, of a large inter-connected system. What none of the approaches reviewed considers explicitly is the long-term, multi-year, environmentally-influenced dynamics of investment programmes, technological improvement and innovation, total system demand and capacity, and demand forecast uncertainty at various periods ahead. Thus, the longer construction period of a larger plant may lead to larger forecasting errors and costs to the total system; no plant level analysis will identify such a drawback. Again, cost escalation affects all plants, and for comparative purposes it is usual to compare plants on a "constant price" basis; but the larger units with longer construction periods may suffer more from cost escalation and interest charges, and it may be wrong to dismiss this effect by price deflation.

A U.K.-based simulation study by Abdulkarim and Lucas [1], based on alternative strategies over the years 1965/66 to 1974/75, concludes:

> "...the economies of scale in very large plant have not been sufficient to offset the attendant disadvantages. Allowance is made for the variation with capacity of the capital cost, thermal efficiency, construction time, planning margin and availability. It is concluded that better results might have been obtained with sets between 200 MW and 300 MW."

The analysis is acknowledged to be not wholly conclusive, but the authors point out:

> "... what the analysis does now show is that there are conditions where economies of scale are outweighed by other factors, that these conditions are not especially remarkable, that they seem to have been satisfied by the CEGB system and that supply units in developing countries, where comparable decisions have now to be taken and where the disadvantages of scale are more pronounced, should examine carefully the case for large generating units in local circumstances."

A rather different study, by Corti [17], compared the performance of the U.K., Electircitè de France, and the Rheinisch-Westfalisches Elektrizitatswerk, in terms of their aggregate performance in three areas: finance; technical performance; and industrial relations. His conclusion was:

> "...the argument for a unitary, concentrated structure for electricity production in advanced industrial countries, resting so heavily on economies of scale, remains a theory only. The past twenty years' experience suggests that advanced industrial countries can have a deconcentrated, devolved system without apparently suffering financial, technical or industrial-relations penalties. In fact the reverse appears to be the case. Evidence does not point to biggest being best."

A wide-ranging critique of both nuclear and coal-based centralized energy production strategies has been made by Lovins [51 & 52], and supported by other advocates of "soft path" technologies, following Schumacher [69]. In addition to social and political factors, which are his main emphases, Lovins claims technical and economic advantages of small scale:

"1) Virtual elimination of the capital costs, operation and maintenance costs, and losses of the distribution infra-structure (see below).

2) Scope for greatly reducing capital cost by mass production if desired.

3) Elimination of direct diseconomies of scale, such as the need for spinning reserve on electrical grids.

4) Major reductions in indirect diseconomies of scale that arise from the long lead times of large systems: for example, exposure to interest and escalation during construction, to mistimed demand forecasts, and to wage pressures by a large number of strongly unionized crafts well aware (as in the Trans-Alaska Pipeline project) of the high cost of delay.
The very conditions that make the indirect diseconomies of large scale important make them hard to quantify. Nonetheless, some utility managers are realizing that interest, escalation, delays owing to greater complexity, and the effects of forecasting errors can make a single large plant of capacity C more costly than N smaller plants of capacity C/N with shorter lead times."

The above very brief review and sampling, of the extensive literature on scale in electricity generation, should serve to indicate the division of opinion which exists.

In considering what type of study IIASA could usefully undertake in the field of electricity generation, the classification by Masud [53] of techniques for expansion planning can be used:

(a) academic: "illustrates certain mathematical or physical concepts, but...would not be used in studying the expansion of a large power system."

(b) conceptual: "illustrating broad concepts for power system expansion. Although it makes many assumptions, the assumptions are consciously made, and the results are useful for planning power systems. The study may or may not reference a particular power system."

(c) screening and (d) reinforcement: "A screening study will be defined as one which does reference a particular system. It makes fewer assumptions than a conceptual study, and is of sufficient detail to yield a few solutions for detailed analysis. These solutions are then refined to include equipment, environmental and socio-political considerations in detail, and ultimately to yield a single solution. This will be defined as the reinforcement study."

The study referred to appears to belong in categories (c) and (d), "screening" and "reinforcement." The IIASA objectives would be on levels (b) "conceptual" and even (a) "academic." Its suitability as a case study would therefore depend on a broadening of the terms of reference beyond the immediate decision. Such broadening would not be in the direction of considering large numbers of other plant expansion alternatives in the country, but should comprise some or all of the following dimensions:

(a) a long historical perspective - e.g. the post-war development of the country's electricity system; and of its largest and average size of units; concentrating on key decisions, the reasons for them, and the subsequent performance;

(b) a similarly long-term (e.g. 30 year) view of the possible futures of the country's electricity supply industry, and of the role of plants like that proposed in this future;

(c) a broader view of the place of industrial centres like that proposed, in the country's economy and society;

(d) an examination of the proposed decision, and the implications of the alternative unit sizes, from several of the disciplinary viewpoints reviewed in Section 4: and a comparison of these disciplinary approaches.

6.3 Diffusion, Barriers to Diffusion, and the Growth of Relevant Contexts

In 5.4, it was suggested that a suitable case problem for research might be directly identifiable from the discussion of new areas of environmental change and related methodological difficulties. In this section, such a possible research area is outlined.

A common structural feature of problems of long-range planning and problems of scale is the way in which successive incremental decisions, individually correct on local criteria, may preclude consideration of strategically sounder decisions; or may defer their consideration or realization until a point is reached where the switch to a sounder strategy would require the abandonment of too large a commitment. In many situations, an apparently correct decision, taking account of the "relevant context," fails to anticipate the way in which the relevant context may change - usually expanding - in later years. This situation is diagrammatically shown in Figure 7, and may be illustrated by many examples, such as those discussed below. The common theme is the role of barriers whose presence or absence, creation or elimination, will facilitate or inhibit the growth of "relevant contexts"; the close relationship between scale of environemnt, and the scale of individual organizations and smaller units, has already been discussed.

The historic tendency has been towards the reduction of barriers to diffusion, of goods and ideas, as transport and communications have been cheapened and simplified. The consequence is a general enlargement of relevant contexts, with many easily observable beneficial effects. However, the existence and exploitation of major economies of scale can itself create barriers, of two sorts. Firstly, a world or a market dominated by large-scale, low cost producers constitutes an obstacle to the survival or introduction of small-scale producers. Secondly, an industry which has invested heavily in capital-intensive facilities will seek to defend its investment by opposing the introduction of new technologies, however potentially advantageous, if these threaten the dominance of the existing organization or the value of its equipment and expertise. The unchecked pursuit of economy of scale may thus tend towards the creation of large, conservative, self-justifying and self-perpetuating establishments, resistant to innovation, intolerant of diversity, and ultimately vulnerable even to minor environmental change.

A possible project would be to examine, quantify, describe and model this process of evolution of scale and reduction of barriers, within any suitably chosen field of industry or similar activity. The examples below illustrate both specific case examples, and some of the potentially relevant disciplines and methodologies.

Example 1: "Keep Left": At some point in the past, it would have been relatively simple and cheap to ensure that Britain's "rule of the road" was the same as that of continental Europe; the longer such a change is deferred, the greater becomes the cost of abandoning the commitment.

Example 2: Currency: Notwithstanding Example 1, the U.K. did decimalize its currency a few years ago; over a century after the advantages of doing so had been pointed out, and accepted in principle by the government.

Example 3: Technical Standardization: The above may be seen as specific examples of the general question of technical standardization, which applies in many fields: radio and electrical equipment, engineering standards in general, railway gauge, road signs, even legislation and language itself.

Technical standardization, like physical connection/separation, may be used either way: the adoption or preservation of incompatible standards preserves barriers and restricts the growth of scale; uniformity of standards reduces barriers and encourages growth of scale.

On a decision relating to technical standardization, such as the British rule of the road, there is relatively little emotional, cultural or aesthetic attachment to a basically arbitrary choice on a purely functional matter. In such dimensions, we have a more purely "technical" problem, which one would expect to resolve by economic criteria. At the other extreme, matters such as regional or national employment patterns, legislation and language may be central to human feelings of identity. In such cases, only extreme economic disadvantage or externally-imposed compulsion will persuade people to abandon their distinctive systems, however small the scale.

Even within some technical fields, there are strong arguments against early standardization backed by large-scale commitment. For when technical progress is rapid, such standardization could act as a brake on progress. It may stifle development, or constrain it to evolutionary development of present technology, even when revolutionary change is possible and desirable: Braun and MacDonald [12] have shown how unlikely the development of the electronic transistor would have been within the large, established manufacturers of vacuum tube valves. This is a field to which technological forecasting might usefully contribute, in assessing when standardization should be encouraged. Fick [28] has outlined a similar problem in the field of computer software, where there appear to have developed structural barriers to the evolution of more efficient languages.

<u>Example 4</u>: <u>Unanticipated Field Effects of Aggregate Behaviour</u>: In a less directly technological dimension, but one requiring the development of technical understanding, there may be unknown, or only partially understood, environmental field effects of scale. Because they are unknown, their emergence typically follows the decisions and commitments giving rise to them. Ecology has yielded many examples. For example, large fields in agriculture yield advantages of mechanization, with higher labour productivity. But the elimination of trees may lead to soil erosion; the elimination of hedges may remove birds which had previously been beneficial in pest control. This is not to say that the larger scale is wrong; but it demonstrates the need for developing a fuller understanding of the system being altered, before change is implemented on an irreversible scale.

Holling [38] has described this process of potential development of an intrinsically unstable system, with illustrative examples from ecology. The classic Huffaker experiment is particularly interesting:

> "...when there was unimpeded movement (of the creatures being studied) throughout the experimental universe (a homogeneous world, therefore), the system was unstable and the populations became extinct. When barriers were introduced to impede dispersal between

parts of the universe, small-scale heterogeneity was introduced and the interaction persisted. Thus populations in one small area that suffered extinction were reestablished by invasion from other populations that happened to be at the peak of their numbers."

Such issues are not necessarily confined to ecology and the natural sciences. A structural feature of larger scale has been the reduction or elimination not only of hedges between fields, but of the barriers of cost, distance and communication difficulty which once separated people, or delayed interactions between them. This "homogenization" leads to accelerated diffusion and standardization not only of technology, but of ideas. Television, universities and plastic toys, steelworks and pharmaceuticals, international institutions and IIASA reports, become standardized in form, technology, appearance and use throughout more and more of the world. This may be a matter for sentimental regret in some minor fields; but has grave risks in areas impinging on human life and well-being. Education, health care, and technologies with significant impact on human behavior or on the living environment, are all areas in which diversity, experimentation, learning and adaptation will continue to be essential. All are areas in which the increase of standardization, scale, and rapid international diffusion will tend to inhibit these essential activities. The process of learning and the control of errors could thus be de-stabilized. A strong parallel can be drawn with the control of epidemic disease, in which one of the most basic needs is to try to restrict movement. The risk is of an "epidemic" of unsound ideas: uncontrolled in its spread, because their unsoundness is slow to become apparent. This is precisely the charge which O'Keefe and Westgate [60] of Bradford have made, in assessing the apparent rise of so-called "natural" disasters. Their argument is that the incidence of natural disasters is attributable partly to the application of inappropriate technology, imported and imposed in standard form, by "experts" who fail to appreciate the innate social wisdom of the local practices. Such practices may have evolved over the centuries in response to real needs of the local situation, but may not withstand the sudden onslaught of large-scale implementation or import of socially inappropriate technology. (See also Sunday Times, 25.6.78, "Disasters: how the helpers make things worse").

The final example again illustrates the application of policies based on defective understanding or models of the dynamics of a situation; again, in situations where scale effects are of major significance.

<u>Example 5</u>: <u>Economic Development and "Protectionism"</u>:

In an undeveloped economy, few manufacturing industries are initially justifiable in terms of "comparative advantage," and in an unrestricted market situation, domestic manufacturing industry will not develop, being uncompetitive with the price of imports. A period of import restriction is required before domestic industry is strong enough to be viable, in its scale or competence.

Unfortunately the imposition of theoretical economic concepts - such as the law of comparative advantage, which is formulated purely in static terms - ignores realities which may display dynamic effects such as technological change, scale effects and learning curve effects. The effects of such policies, as imposed by the International Monetary Fund, have been the subject of growing criticisms (e.g. Peyer [63]). Vietorisz [72] has documented this process with examples (the electric motor industry, in Mexico), and has extended this to a thorough and convincing analysis of the dynamics of economic development, combined with a strong indictment of "comparative advantage."

This argument is not restricted, however, to the economic development of poor countries: the same or closely related issues are raised by Simmonds' analysis of scale in the U.S. and Canadian chemical industries; by Godley [35] and his colleagues, in their argument for selective import controls in the U.K.; by the current GATT negotiations [23]; and by the arguments about tariffs, subsidies, and industrial rationalization in the EEC [22 & 24]. The global significance of trade barriers is highlighted by Roberts [64] and the SARU global model, in which population and the mean value of the trade bias matrix are identified as critical parameters for the avoidance of catastrophe. Although a reduction of the mean trade bias is the preferable direction in the aggregate runs, some "experimental runs of SARUM concerned with raising trade barriers in order for Africa to secure an improved ultimate position are a justification for querying the orthodox free-trade-is-good advice." These runs are reported by Parker and Raftery [62].

These examples illustrate several of the environmental structures referred to by Emery and Trist. They include, deliberately, both "neutral" or "technical" examples, and examples of potential or current controversy. Questions of scale occur at the highest political level, where there may be conflict between efficiency (lowest cost) and self-sufficiency (control) in key strategic commodities. The rationale for Britain's accession to the European Community was largely argued in terms of scale effects; but the principles of comparative advantage, with free movement of capital and labour, and consequent regional or national specialization of role are hardly acceptable in the short term to those local industries facing elimination through "rationalization"; particularly where "local" in the European context means "national."

In the Soviet Union, it is accepted that the scale advantages of the national economic efficiency viewpoint take precedence over the narrower view of any business efficiency [44].

But to return to the practicalities of IIASA's research programme, a project would be required in which the general structural features of barriers and contexts described above could be investigated in the context of a specific industry. An industry of sufficient size and significance to be globally significant would also be likely to provide much of the data from existing published sources. The steel industry, or certain sectors of chemicals, could be appropriate: I.C.I. might be asked about their thinking on "world-scale plants." Some of the rationalization problems being faced by the EEC Industry Commissioner, could also be of interest for study, with reference to the questions of scale involved.

6.4 The Creation, Expansion or Reconstruction of an Industry

In 5.4, the final suggestion for a study on (or significantly involving) problems of scale, was that a project should be sought which would match the skills and interests of existing staff at MMT. In view of their applied, industrial experience, the above title indicates possible suitable areas.

The creation of an industry may mean either the introduction of known technology in an underdeveloped country (e.g. its first cement works), or the attempt of a developed country to develop capability in some new area of high technology - e.g. Britain's development of microprocessors. Expansion is the more conventional area of planning additions and replacements to an existing industry. Reconstruction is applicable to a long-established industry, in which the pattern of products, production facilities etc., may have become increasingly inappropriate because of environmental changes.

References

[1] Abdulkarim, A.J., and Lucas, N.J.D., Economies of Scale in Electricity Generation in the United Kingdom, Energy Research, 1, 223-231 (1977).

[2] Ackoff, R.L., and Emery, F.E., On Purposeful Systems, Tavistock, London, 1973.

[3] Adelson, R.M., D.C.F. - The Other Point of View.

[4] Allport, F.H., Theories of Perception and the Concept of Structure, John Wiley and Sons, New York, 1955.

[5] Ansoff, H.I., Corporate Strategy: an Analytic Approach to Business Policy for Growth and Expansion, McGraw Hill, New York, 1965.

[6] Ball, D.F., and Pearson, A.W., The Future Size of Process Plant, Long Range Planning, August 1976.

[7] Berry, Ya (Ed.), Planning a Socialist Economy, Vols. 1 & 2, Progress Publishers, Moscow, 1977.

[8] von Bertalanffy, L., General System Theory, Foundations, Development, Applications, first published in the U.S.A., 1968, Penguin University Books, 1973.

[9] Bolotryi, K.A., Itin L.I., *Planirovanie optimalnogo razmera predpriyatiya* (Planning of optimal size of enterprise), Machinostroenie, Moscow, 1976.

[10] Booth, E.S., and Dore, J.W.H., The development of large electricity generating units, The World Power Conference, Madrid, 1960.

[11] Boulding, K.E., The Organizational Revolution, Harper and Row, New York, 1953.

[12] Braun, E., and MacDonald, S., Revolution in Miniature: The history and impact of semiconductor electronics, Cambridge University Press, 1978.

[13] Brown, F.H.S., and Booth, E.S., "Coal and oil fired power stations, The 9th British Electrical Power Convention, 1957.

[14] Cameron, D., Three Simple Steps to Determine Optimum Plant Capacity, Long Range Planning, February 1974.

[15] Chandler, A.D., Strategy and Structure, Garden City, N.Y., Doubleday, 1962.

[16] Coleman, A., Land-Use Planning in a World of Shrinking Resources, Long Range Planning, 11, February 1978.

[17] Corti, G., Electricity Industry and Problems of Size, National Westminster Bank Quarterly Review, Aug. 1976.

[18] Dahl, R.A., and Tufte, R., Size and Democracy, Stanford University Press, 1974.

[19] Delombre, J., and Bruzelius, B., Importance of Relative Market Share in Strategic Planning - A Case Study, Long Range Planning, 10, August 1977.

[20] Dewar, R., and Hage, J., Size, Technology, Complexity and Structural Differentiation: Toward a Theoretical Synthesis, Administrative Science Quarterly, 23, 1, March 1978.

[21] Donaldson, L., Woodward, Technology, Organizational Structure and Performance - A Critique of the Universal Generalisation, The Journal of Management Studies, October 1976.

[22] The Economist, A European selling price system for chemicals, 24 June 1978.

[23] The Economist, GATT's trade talks, 17 June 1978.

[24] The Economist, The Politics of Protection, 24 June 1978.

[25] Eilon, S, Structural Determinism, editorial in Omega, 5, 5, 1977.

[26] Emery, F.E., and Trist, E.L., The Causal Texture of Organisational Environments, Human Relations, 18 (1975), reprinted in ref. 27.

[27] Emery, F.E. (ed.), Systems Thinking, Penguin Modern Management Readings, 1969.

[28] Fick, G., Some Comments on the Computer Penetration Study, notes at MMT, IIASA, 9.6.78.

[29] Fisher, J.C., Energy Crises in Perspective, Wiley, New York, 1974.

[30] Forrester, J.W., Urban Dynamics, M.I.T. Press, Cambridge, 1969.

[31] Glagolev, V.N., *Mechanizacija i avtomatizacija truda v promyshlennosti Litovskoi SSR* (Labour mechanization and automation in the Lithuanian SSR's industry), Mintis, Vilnius, 1975.

[32] Glagolev, V.N., *Mechanizm planirovanija i upravlenija programmoi regionalnogo zazvitija TVA* (Mechanism of planning and management of program of regional development - TVA) Proceedings of All-Union seminar "Problems of interbranch management of social production," Moscow, 1977.

[33] Glagolev, V.N., *O mezotraslevoi territorialnoi programme industrializaciji vspomogatelnogo proizvodsva* (On the problem of territorial program of industrialization of auxiliary production), Proceedings of All-Union seminar "Problems of interbranch management of social production," Moscow, 1977.

[34] Glagolev, V.N., *Problemy povyshenija effektivnosti mechanizaciji i avtomatizaciji v promyshlennosti. Tesisy dokladov vsesojuzhoi nauchuoi konfezenciji* (Problems of raising mechanization and automation efficiency in the industry), Proceedings of All-Union scientific conference, USSR's Academy of Sciences publishers, Moscow, 1973.

[35] Godley, W., and colleagues, Cambridge Economic Policy Review, 1976, 1977, 1978; esp. 1976 Chap. 4, The relative decline of the UK manufacturing sector.

[36] Gold, B.,Evaluating Scale Economics: The Case of Japanese Blast Furnaces, Journal of Industrial Economics, September 1974.

[37] Gold, B., (ed.), Technological Change: Economics, Management and Environment, Pergamon Press, 1975.

[38] Holling, C.S., Myths of Ecology and Energy, in ref. [58].

[39] Huettner, D.A., Scale, Costs and Environmental Pressures, in ref. [37].

[40] Huettner, D., Scale, Costs and Environmental Pressures, in ref. [37].

[41] Ickovich, I.A., *Analiz lineino-ekonomicheskih matematicheskich modelei* (Analysis of linear-economic mathematical models), Nauka, Hovosibirsk, 1976.

[42] Imperial Chemical Industries Ltd., Assessing Projects, Methuen & Co.Ltd., London, 1970.

[43] Jantsch, E. (ed.), Perspectives of Planning, OECD, Paris, 1969.

[44] Kamenitsez, S., The Experience of Industrial Management in the Soviet Union, Progress Publishers, 1975.

[45] Kantorovich, L.V., *Ekonomicheskij raschiot nailuchshego ispolrovanija resursov* (Economic calculation of the best use of resources), Academy of Sciences publishers, Moscow, 1959.

[46] Knop, H., and Straszak, A. (ed.), The Bratsk-Ilimsk Territorial Production Complex: A Field Study Report, IIASA, Laxenburg, Austria, May 1978.

[47] Kosygin, A.N., Guidelines for the Development of the National Economy of the USSR for 1976-1980, Moscow, 1976.

[48] Landon, J.H., Changing Technology and Optimal Industrial Structure, in ref. [37].

[49] Lee, T.H., The Case for Evolutionary Optimization, in ref. [58].

[50] Leschiner, P.E., *Obosnovanije omtimalnych razmerov predpriyatij v promyshlennosti* (Substantiation of optimal sizes of enterprises in industry), Institut narodnogo choziaistva im. Plechanova V.G., Moscow, 1970.

[51] Lovins, A.B., Energy Strategy: The Road Not Taken?, in ref. [58].

[52] Lovins, A.B., Scale, Centralization and Electrification in Energy Systems, in ref. [58].

[53] Masud, E., Advanced Concepts for Bulk Power System Studies, Long Range Planning, 10, October 1977.

[54] Meyer, R.F., On the Relationship among the Utility of Investment Strategy in an Uncertain but Time-Invariant World, Proceedings of the Fifth International Conference in Operational Reseach, Tavistock Publications, 1970.

[55] Nekraso, N.N., *Problemy teorii i praktiki razmeshchenija proizvoditelnych sil SSSR* (Problems of theory and practice for distribution of productive forces of the USSR), Nauka, Moscow, 1976.

[56] Nerlove, M., Estimation of a Cost Function: Returns to Scale in Electricity Generation, in ref. [57].

[57] de Neufville, R., and Marks, D.H., Systems Planning and Design: Case Studies in Modeling, Optimization and Evaluation, Prentice-Hall, 1974.

[58] Oak Ridge Associated Universities, Future Strategies for Energy Development: A Question of Scale, Oak Ridge Associated Universities, 1977.

[59] OECD, Transfer of Technology for Small Industries, OECD, Paris, 1974.

[60] O'Keefe, P., and Westgate, K., Preventive Planning for Disasters, Long Range Planning, 10, 3, June 1977.

[61] Ozbekhan, H., Toward a General Theory of Planning, in ref. [43].

[62] Parker, K.T., and Raftery, J., The Sarum Global Model and its Application to Problems of Interest to Developing Countries, proceedings of the International Conference on Systems Modelling in Developing Countries, Asian Institute of Technology, Bangkok, 8-11 May 1978.

[63] Peyer, C. The Debt Trap: The IMF and the Third World, Pelican, 1974.

[64] Roberts, P.C., SARUM 76 - A Global Modelling Project, Futures, Feb. 1977.

[65] Sahal, D., A Theory of Progress Functions, discussion paper 78-10, Wissenschaftszentrum, Berlin, 1978.

[66] SARU (Systems Analysis Research Unit), SARUM 1976: Global Modelling Project (Section to 7.3), Depts. of the Environment and Transport, London, 1977.

[67] Schneider, K.R., Planning Without Vision - Roots of the Urban Disaster, Long Range Planning, 10, June 1977.

[68] Schroeder, T.W., and Wilson, G.P., Economic Selection of Generating Capacity Additions, Transactions of the AIEE, 1133-1145, 1958.

[69] Schumacher, E.F., Small is Beautiful, Harper & Row, New York, 1973.

[70] Simmonds, W.H.C., The Canada : U.S. Scale Problem, Chemistry in Canada, October 1969.

[71] Simmonds, W.H.C., Stepwise Expansion and Profitability, Chemistry in Canada, September 1969.

[72] Vietorisz, T., Diversification, Linkage, and Integration Focus in the Technology Policies of Developing Countries, in ref. [59].

[73] *Voprosy optimalnogo razmera predpriyatij v promyshlennosti SSSR*, (Problems of optimal size of enterprises in the USSR industry), Nauka, Moscow, 1968.

[74] Weinberg, G.M., An Introduction to General Systems Thinking, Wiley, New York, 1975.

[75] Winger, A.R., Urban Economics: An Introduction, Charles E. Merrill, 1977.

[76] Woodward, J., Industrial Organization, Oxford University Press, London, 1965.