

# RESEARCH PROGRAM EFFECTIVENESS

*Proceedings of the Conference  
Sponsored by the Office of Naval Research  
Washington, D. C., July 27-29, 1965*

*Edited by M. C. YOVITS  
D. M. GILFORD  
R. H. WILCOX  
E. STAVELEY  
and H. D. LERNER*

---

**GORDON AND BREACH, Science Publishers, Inc.**  
NEW YORK • LONDON • PARIS

## Chapter X

### A Graph Oriented Model for Research Management\*

O. Morgenstern  
*Princeton University*  
*Princeton, New Jersey*

and

R.W. Shephard  
*University of California*  
*Berkeley, California*

and

H. Grabowski  
*Princeton University*  
*Princeton, New Jersey*

---

\*Based on a report by Mathematica to Office of Naval Research under Contract Nonr-4552(00), Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

## ABSTRACT

The fundamental decisions of research management, such as choosing the research to support, deciding at what level it should be supported, and measuring the results of such support, are most often made by largely intuitive procedures. This paper offers some formal concepts and methods designed to illuminate the underlying structure of research process. A key feature of this structure is the set of manifold interconnections and feedbacks between different research activities and projects. In the paper the various interrelations are displayed by a graph, or "tree," to see how sponsorship of research is related to some government objective in both a direct and indirect fashion. Frequently one might expect to find that the indirect or "derived" effects will be more significant than the direct effects. The interrelationships that characterize the structure of research are also expressed by means of a matrix which has a mathematical correspondence to the "tree." Such a matrix allows one to estimate the derived effects of research in a quantitative fashion if one is able to specify coefficients of research transfer between the various research activities. It happens that the matrix equation system which relates the total (direct plus "derived") effects of research to the initial direct inputs is similar in form to the Leontief input-output analysis. In our paper several normative models are constructed which use Leontief systems to characterize the structure of the research process. These models are only meant to be suggestive of the possible uses of our concepts and methods. Our major concern in this paper is with formally characterizing the actual underlying structure of the research process so that the resulting concepts can be used as a supplement to more intuitive procedures based on the research administrator's experience and intelligence in whatever manner he deems desirable.

The second half of the 20th Century has been characterized by some as the beginning of the "Research Age." It is indeed true that in the last decade and a half, organized research has undergone an explosive development and that the welfare and security of our nation is increasingly dependent upon the proper management and planning of our research activities. Management of research remains, however, a task of extraordinary complexity. It is essentially an unstructured problem, dealt with largely on an ad hoc basis.

This statement is true in spite of the fact that organization of research goes back at least to Aristotle, who received large grants from Alexander the Great, with which he financed and organized expeditions of researchers to provide him with descriptive material which he incorporated into his books. Centuries later, Bacon of Verulam in his *Novum Organum* (1620) stated that the three most important discoveries up to then--compass, printing, and gunpowder--had been made by accident and that incredibly more could be produced by systematic procedure. He conceived the idea of truly scientific laboratories; yet it was not until the 19th Century that these arose. Now, almost 100 years later, when we have an abundance of such institutions, we still lack a body of accepted rules which would securely guide those whose task it is to determine what funds to raise for research, how to distribute them among conflicting demands, and how to evaluate the results of the research activities.

The problem of research management is not new. Everyone in control of some funds and facilities for research has encountered it, given thought to the problem, and had to make decisions as best he could. These ad hoc decisions lack the firm foundations of a theory from which rules of behavior could be derived. It is not surprising that such theory does not now exist since there are still great difficulties in establishing valid theory in those closely related fields where preliminary progress would have to be achieved, especially in the field of public investment. Thus while a workable theory of research management is the ultimate aim, we are far from achieving it at present.

Another way of looking at research processes is to view them as part of a game played by the researcher with nature, where the usual assumption is that nature is merely indifferent about being found with one another and with the supporting agencies--clearly a game of very great complexity. This line of thought shall not be followed explicitly in the subsequent considerations. However, it should be pointed out that one of the main difficulties is that one will have to start with a finite system (whatever its characteristics), but that one is really confronted with an infinite system. Concluding from a finite to an infinite system is one of the severest difficulties encountered in scientific methodology.

With this restriction in mind, it will nevertheless be shown that some fairly rigorous models can be formulated which will at least provide a set of conceptual tools with which further descriptive work can be carried out, all to be done in the expectation that from a variety of such procedures some useful rules may emerge.

What a model is to be in this context is not a simple matter. We shall be satisfied if we can at least describe possible relationships, without necessarily demanding that they be true or empirically completely valid ones. Only after the possibility of a model has been established can the further questions be approached. In this paper we shall be content if the first question can be answered in the affirmative (as we believe to be the case). A description of some elements of a model will, of course, not involve the establishment of an axiomatic system, which would normally be the ideal test of a model. The above remarks concerning the need to conclude to an infinite system make this fact amply clear.

## PRESENTATION OF MODEL IN GRAPHICAL FORM

### Allocation Problem

In this paper we will be considering the problems of research management which confront large government or non-profit agencies supporting research in a wide variety of different fields. The examples of our analysis will therefore be concerned with research policies designed to fulfill broad government or social objectives.

This, of course, does not mean that the methods and techniques to be developed do not have applicability to research programs that are quite different in scope or purpose. However, we shall deal here specifically only with their relevance to the problems of the large non-profit research supporting agency.

The basic problem facing an agency that is set up to support research at various levels is clearly the following: given a certain amount of money, how is this amount to be distributed optimally among different possible uses when there are more of them than can be satisfied with the available funds? This looks like an ordinary economic problem; and, as will be shown, it has many similarities with one. There are formidable complications which set it apart, however.

No rational allocation is possible among research fields, unless there exist clear notions about values, preferences, and utilities. Alternative uses of resources must exist and must be compared with one another. This comparison requires the introduction of a notion of cost which, in conformity with modern economics, has to be viewed in terms of "lost opportunity." That is, the support of one research project is the inability at the same time and with the same funds to support another project. This is the true "cost" of supporting the chosen project. To justify the choice made, the supported project must be more important, promising, or valuable than the one discarded. This is, of course, what every support agency tries to establish, even though a formal calculus for determining values may not exist. In that case, rules of thumb are used, experience is called upon, hunches play a role, etc. The report on which this paper is based includes an attempt to indicate how the organization of the support agency may bring about the best approach to the use of these qualitative and subsidiary elements in the decision process.

In ordinary economic allocation, there is a comparison of a known utility with another one. That is to say, we know what an automobile will do, or the consumption of a pound of meat, sugar, etc. We form an *expectation* of the effects, based on experience or at least on a projection of effects from similar situations in which we have encountered the object of our choice. If there were no experience of this kind, no preferences at all could be formed; and rational action, as understood commonly, would be impossible. Not being sure of the effect upon us, if the event is realized, necessitates

the introduction of *expected outcomes* with certain probabilities.

In research the complication is as follows: a decision of allocation of the right amount is to be made when there is sometimes no information as to what effect, if any, can be expected. One can compare, *ex post*, a phenomenon or result of an allocation with others; but there is generally no information about what would have happened if another allocation had been chosen. This holds *ex ante* as well as *ex post*, since no event has occurred in the unsupported area, precisely because it was not supported, though the suspicion remains that, had support been given, something of possibly greater value might have materialized. So we compare the result  $X^1$  (realized) at  $t_1$  with our hoped for  $X$  at  $t_{1-n}$  (when the allocation was made) and with the unknown effects  $Y$  at  $t_1$ , about which we can only form vague ideas, since they remain at  $t_1$  just as inaccessible as they were at  $t_{1-n}$ . Thus the difficulty is that only realized effects can be compared with one another, allowing only a partial evaluation of the allocations made. But in research the essence is that an unknown effect was aimed for, which in most cases can only be imperfectly described and in some cases, not at all. The more basic the research, the more is this the case. In engineering research, concerned with a novel application in a well defined and well controlled environment, most of these difficulties are diminished.

It follows that the agencies supporting basic research encounter greater uncertainties than those primarily involved in research directly connected with improvements of known devices, methods, etc. Most agencies support both basic and applied work and therefore have the problem of allocating funds among *areas* and then of choosing optimally within each area.

A further complication is the time factor. Expected effects may happen, if they do at all, at times quite different from those anticipated. This fact affects the value of the results, one achieved early normally being more valuable than one obtained later. Yet some results are interdependent and the exploitation of an early outcome may have to wait for another one, whereby its value is temporarily diminished. This intertwining of research activities, which extends from one to the other and variously over time, is one of the basic characteristics of the field and represents one of the most formidable difficulties of analysis. Furthermore, whatever patterns can be

discovered are not likely to remain stable for long periods, because instability is precisely the consequence of new effects produced by the entire research effort. It is natural, however, to look at a stationary state first, in order to see whether any kind of structure can be discovered at all.

In order to approach the problem of allocation of financial funds and human effort, it is necessary to obtain a description of the entire process of research activities as they go on at various levels and in diverse fields. If a model can be constructed at all, it would be the first indispensable step towards an operationally useful procedure for any research-supporting agency to discover its own position in the entire complex of research activities. Clearly, no agency can support all fields, yet it is fully exposed to the effects of research carried on by otherwise supported activities, on which in turn it also produces some feedbacks.

### The Graphical Model

*Classification Schemes.* We turn to a discussion of our model of the structure of research for planning and management. As stated above, the structure of research is similar in form to other activities in the economy; it involves the utilization of scarce resources as inputs in order to obtain desired goods or services as outputs. Research outputs, however, are not measured or evaluated like commodities in the market. Clearly, they are multi-form and multi-facet, qualitatively complex, and not at present measurable or describable in simple terms; they range from mathematical formulations of phenomena to critical experiments and laboratory devices. Published papers are descriptive of work done but, in themselves, as so many pages of printed material, cannot serve to measure outputs of research, which are ideas, conceptions, theories, models, experimental results, devices, new materials, et cetera.

The value of the outputs from any given research project is intricately intertwined with the outputs of the projects. The accumulation of knowledge is a self-generative process. One idea depends on a host of others and leads to another in an endless chain. In trying to evaluate the outputs of research activities, it is thus more significant to seek the consequences for further research in the given and related areas of research than to try to measure outputs as values of an exchange economy. When results of research can be

used directly to resolve or facilitate a particular social purpose, government responsibility, or national objective, a social utility can possibly be identified; but the full social value is not restricted to such direct impacts, which may be purely temporary. A more obscure result may lead indirectly but surely to other research outputs, which have greater social impact. Thus the interrelationships between research activities leading to derived effects are essential for any true model to serve the planning and management of research. It is therefore clear that the worth of any given research support should be determined from both derived and immediate effects. In fact, for basic and applied research not directed to obtain particular social ends, practically all of the social worth comes from derived effects.

While it is an extremely difficult task in actual practice to obtain a quantitative measure or index of the expected interrelationships between research activities, one can usually specify at least which activities are interconnected and which are likely to be independent of each other, for all practical purposes, over the next decade. As a first step, one can then utilize this information to build a graphical model of the type which we suggest below. Such a graph may be used to make qualitative conclusions about questions of research management. As methods and procedures are developed which are capable of assigning meaningful numerical estimates to these interrelationships, one can formulate the graphical model into a matrix and use it to arrive at quantitative conclusions about different research allocations.

In order to construct our graphical model, one must first make a reasonable comprehensive taxonomy of basic and applied research fields. For this purpose, the classification system of the National Science Foundation may be used, supplemented by project statements for the more directed research activities carried out in government laboratories. Development activities should not be described in the same model with research since they have, presumably, specific objectives which can be programmed to yield well defined systems.

The specification of research activities is not something which can be expected to result in a unique classification system. In each field of research, the expert can conceive of finer distinctions which lead to more and more categories of research; and one must

proceed in some fashion without resorting to the extreme of taking each researcher as involved in a separate distinct activity. Thus, we are involved in a procedure of aggregation; and the guidelines for the choices involved should be drawn with an eye to the uses contemplated for the classification system.

*Structure Modeled by Linear Graph.* Once we have constructed a classification system, we may display on a linear graph the interrelationships between research activities and connections to government activities, as illustrated in Figure 1. The *nodes* of this graph represent distinctly identified research activities, arranged so as to progress from Pure or Basic to Applied to Direct research, and the *directed arcs* show the interrelationships.

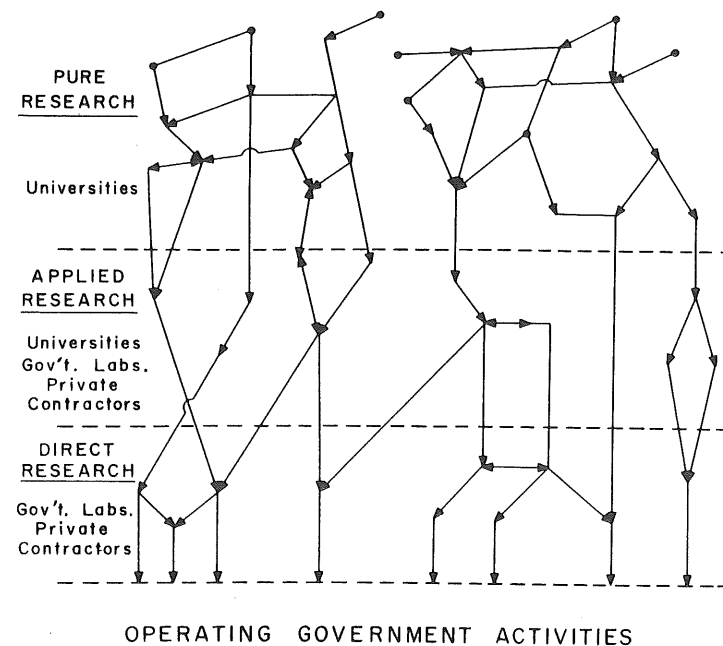


Figure 1. Interrelationships between Research Activities

Each *node* corresponds to an aggregate of research efforts in some category of the taxonomy of basic and applied research fields, involving groups of people ranging from a single individual to large groups scattered among many universities or concentrated in large laboratories. The number and complexity of researchers involved depend upon the degree of aggregation, which may vary as one progresses from Basic to Direct research.

Direct research is distinguished from Applied research as primarily bearing directly upon some specific problem area of government activities, while Applied research generally deals with a larger class of problems having identifiable potential applications in several Direct research areas and even, at the same time having direct use for the resolution of some governmental problems. The distinction is as much one of motive as it is one of the result likely to be obtained. A direct research effort may sometimes lead to generally formulated solutions and methods which have application and use in apparently dissimilar areas of other direct research efforts. They thereby have the property of Applied research of being applicable to several distinct Direct research areas. Pure research is undertaken to gain new knowledge in itself and is not consciously concerned with specific practical problem areas, although there may be an awareness that the results may be significant or important for a class of engineering problems, such as purely mathematical studies were for propellor design. Here again, the distinction between Pure and Applied research is mainly one of motive; and some work in the latter level of research may develop into a pursuit of knowledge for itself, stimulated by a class of applications to be made.

For all three levels of research effort, the distinctions depend upon the degree and kind of aggregations made in defining research activities. As in all aggregations, there is a certain freedom or arbitrariness regarding the amount of lumping together of different activities or processes.

The *arcs* between nodes may be uni- or bi-directional, indicating independence of one of the two fields from the other or mutual dependence. Those applied research activities which bear directly upon specific problem areas of government responsibility will be primarily uni-directional toward a platform of government responsibility, having the appearance of the trunks of treelike graphs

whose upper stems and branches fan out as one proceeds to more pure research activities, with greater interconnections between nodes, i.e., more like a web than a tree.

When a particular research activity is regarded in its relationship to some operating government activity, very often one may find no single, direct arc of connection to it and be inclined to regard the research activity as having no bearing upon the problems confronted. Yet, study of the graph will show relevance through one, or more, other research activities and thus indirectly that it has strong influence. The test for relevance is whether there exist connected paths between the given research activity and the particular government problem studied. Within the accuracy of the graph, one may seek research activities to support which have bearing, direct and indirect, upon some broad area of government responsibility. It is clear that if a direct arc exists, the connection is strong and intuitively convincing. If there are many nodes through which the connection has to pass, the matter is more complicated. The significance of research activity at the most distant node for the government problem arises from the derived effects along the totality of connected paths to the problem area in question. These derived effects are not simple to perceive since they interact and are carried along with the effects of the research activities at all of the nodes of the connected paths. More will be said about these derived effects when a quantitative form of the graph model is considered in the section on "Presentation of Model in Matrix Form."

*The Three Research Trees.* There are three possible variants of the linear graph described above which will be useful to the research supporting agency in making allocations. We shall denote these three graphs as the *Fundamental tree*, the *Relevance tree*, and the *Support tree*. We shall now discuss each of these concepts in some detail.

(a) The *Fundamental tree* describes the "objective" relations among nodes, i. e., those influences of each node on others as can be established from our understanding of the interconnectedness of science and engineering. It shows necessary and possible connections (at the given level of aggregation), such as that between the mathematical theory of differential equations and the theory of heat transfer and related problems in engine construction, or between any mathematical theory and the various theories of physical

phenomena. There is a web or circularity within the field of pure research, a lesser web and more direction in the applied research area, and usually a one-directedness in the final phases of direct research. But given a certain state of knowledge in pure science and technology, and neglecting time, the relationships between the aggregated nodes of the Fundamental tree are reasonably objective (i.e., they are reconstructable by different individuals with knowledge of science and engineering). This tree tells us nothing about financial support; it exists whatever the level of support may be.

(b) The second concept pertinent to research management is the "*Relevance tree*." This is a fully connected sub-graph of the Fundamental tree which delineates, again *objectively*, the *specific* influences leading out of the pure research via applied research to the specific organizational purposes and hardware programs of the research supporting agency under consideration. This is one interpretation of the tree described in Figure 1. If the organization has a very narrowly defined or limited set of objectives, the tree will be a very small and slender sub-graph (though extending into the fundamental research area). If it has very ambitious goals, hoping to utilize the fruits of wide areas of industrial and government research, the sub-graph will be broader; but it will never coincide with the whole Fundamental tree, which describes all research going on in the country, some of which will never have any perceptible connection to any specific agency. A complete coincidence would be a sign of an absurdly large aggregation.

The Relevance tree as such tells us immediately nothing about support, actual or desirable. It merely informs us about the specific relations and subsets of the Fundamental tree which are of concern to the research supporting agency and the other agencies; and it shows where these agencies have common interests and where they begin to diverge. Once the shape of this tree has been determined, it is possible to begin to evaluate the effect of research interrelationships on organizational *goals*. The various arcs are necessary, considering our understanding of these relationships in an essentially stationary state. An examination of the graph would show which connections are relevant to the achievement of some organizational objective. If in actuality there is a lack of performance at one or the other node (a fact to be determined by records, evidence obtained from scientists, et cetera), then a reason is given for

support of the node at which the arc originates in order to stimulate the desired flow in the next or at any stated later period of time.

(c) The above considerations lead us to the formulation of our final graph, the *Support tree*, which need not be a fully connected subtree of the *Relevance tree*. Rather, it is a picture superimposed on the Relevance tree of research activities supported (nodes) and arcs of influence, which the agency thereby hopes to generate or has generated and tries to keep going or to strengthen.

Support of research is reflected at the nodes. If it is undirected (i.e., without prescription as to the kind of results sought), the aim of the support is general enhancement of the generation of knowledge, and the benefits to the organization arise from the diffusion of this knowledge to other related research activities (i.e., undirected derived effects). In such support, new nodes of research activity may be supported by the agency. If the research support is intended to yield results which affect and facilitate other specific research activities, the arcs of intended effects on other nodes, along with the node supported, need to be distinguished. Such directionally effective research support will typically be attempted in the Applied and Direct research areas, near the final engineering stages. In these cases, there is *strict guidance* of the research. Specific instructions are given concerning the direction in which one should look, prescriptions and projects of very definite character are worked out, quite different from the support of a node in the area of basic research, where one will gladly accept any connection to another node that may arise. It may, of course, be that no new arc forms, no additional influence on existing arcs is observed, though the node itself may benefit. The uninstructed support will largely be in the area of fundamental research; the nearer we come to final, practical applications the more definite will the instructions become that the supporting agency designs. Even here there are dangers, since creativity is just as much present in applied fields as in fundamental research; and many ingenious turns cannot be foreseen. Indeed, results in the final technical area may produce, on occasion, even an arc back to the area of fundamental research, for example, by creating the possibility of new measurement tools which are required to provide new inputs, even for some of the most rarefied fields of theory. The history of mathematics is replete with cases in which new problems were posed by physics, problems which arose only because new experiments had become possible.

An actual Support tree may look something like this

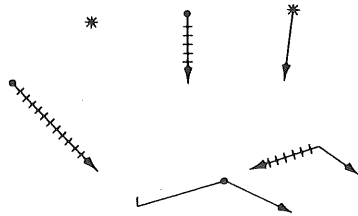


Figure 2. Support Tree

In Figure 2 a segmented arrow indicates general support in the hoped-for direction, while a plain arrow indicates that the support is tied to a directive, for the node, to try to evolve results that will augment the state of activity at a specific subsequent node, etc. These arrows are superimposed on the Relevance tree. An asterisk at a node indicates that it is being supported but that, at best, only an expectation exists that consequences of the support will affect other unspecified nodes. Clearly, these do not have to be supported everywhere, since there are also other sources of support, although the agency may very well want to add to those expected streams in order to obtain a larger or speedier effect.

At each node, the dollar value of support may be entered; it will have to be compared with a *value index* for the ultimate results expected, i.e., payoff for organizational responsibilities.

At present, one is far removed from possessing a generally acceptable method for assigning numerical values for different expected research results. Even *ex post* values can only be determined roughly by stating orders of magnitude. Nevertheless, it is helpful to describe the research process in terms of the above analysis, since understanding of a morphological kind must precede a deeper analysis.

An important lesson to be learned from the history of science is that, although it is an easy thing to do, there is no point in demanding too much. Progress has always been connected with limiting objectives in a modest way. In this specific case, one should recall that it has taken generations to clarify the economic concept of utility and preference and that this work is not concluded. Yet economic utility is a far simpler notion than the utility of new scientific or engineering discoveries, especially when the discoveries have multiple uses and ramifications.

In the next section it will be shown that it is possible to distill from the graphs described so far a considerable amount of the information which should prove to be useful in applications. The procedure has the advantage of making use of some recent advances in economics, combined with proper mathematical techniques.

#### PRESENTATION OF MODEL IN MATRIX FORM

As a model of the structure of research, a linear graph provides several insights which are not easily obtainable from isolated evaluation of research programs. Such a graph makes possible an analysis of the interrelationships between research projects, at least in qualitative fashion. For a more sophisticated treatment of the derived effects of research, however, we must introduce transfer coefficients which define that fraction of the research in one activity carried over to another with which it is connected. Once we have introduced such coefficients into our analysis, we may use the mathematical methods and techniques associated with matrix algebra to make quantitative inferences concerning the derived effects of research activities. We will proceed by first treating the research process as timeless in order to establish the steady state implications of our assumptions. Later, we shall introduce some assumptions concerning the time nature of the interrelationships. Although the actual coefficients may not be known or determinable with great accuracy, we may, for the present model purposes, assume their existence.

We now proceed to introduce these coefficients into our analysis. Suppose that for any two research activities  $A_i$  and  $A_j$ , the transfer coefficient  $c_{ij}$  defines the transfer of research results of  $A_i$  and  $A_j$ . In general, we assume these coefficients to have the following properties:

- (1)  $0 \leq c_{ij} \leq 1$
- (2)  $c_{ij} = 0$  if no arc connects  $A_i$  in the direction of  $A_j$ , otherwise  $c_{ij} > 0$
- (3)  $c_{ij}$  is not necessarily equal to  $c_{ji}$
- (4)  $c_{ii} = 0$

Property (1) expresses that the transfer coefficient represents a fractional or percentage transfer of research effort from one research activity to another, which we assume cannot be negative and can at most have a value corresponding to a transfer of 100%. Property (2) relates to the one-to-one correspondence between the graph and the coefficients (unitless numbers), which represent arcs on the graph. Property (3) states that the interchange between two activities is not necessarily equal in both directions; and Property (4) states that the derived effect of a research activity for itself is zero. This statement does not mean that research in a given field has no effect within this field; this effect will be treated as an initial one in the analysis below.

Assuming the above properties of the transfer coefficients is tantamount to assuming that the functional form of the interactions between research fields is linear in nature. This, of course, is a first approximation to the actual structure of any system of research activities. From an empirical standpoint, it may be either a good or bad approximation. One can be confident that there is a positive relationship between different research activities, but whether this positive relationship is best described by a linear function or one that exhibits increasing or diminishing returns is an open question and one on which research would certainly be welcome. The whole problem of obtaining an adequate index for measuring research outputs which is a prerequisite for answering this question is certainly a central problem of research management at the present time. The linearity assumption does have the advantage that it is much easier to implement in practice than assumptions involving nonlinear or stochastic elements. The significance of this fact will become apparent when the input-output structure arising out of the above assumptions is discussed and when it is shown that inputs and outputs can all be expressed in similar units. The present state

of empirical knowledge on the structure of research does, however, make it highly desirable that the implications of other types of functional interactions be analyzed and compared with the linear case. Nevertheless, the scope of the present paper will be an analysis only of linear interactions between research activities. Hopefully, at some later time we will be able to supplement this case with others involving various nonlinearities.

For a system of research activities interconnected in a manner described by transfer coefficients such as those above, we may trace the effects of any research effort in the following manner. If a given level of research effort is being performed at research activity,  $i$ , say  $y_i$  man-hours (we shall use man-hours as an index of effort for the present), then this implies that there will be a "derived effect" or transfer of research effort to the  $j$  activity equal to  $y_i c_{ij}$ . By a derived effect here, we mean that  $y_i$  man-hours of work on activity  $i$  has an effect on activity  $y$  which is equivalent to  $y_i c_{ij}$  man-hours of work being directly done on project  $j$ . The derived effect at research activity  $j$ , however, will have its own impacts on other fields of research, i.e., will produce a second stage of derived effect; and these secondary derived effects will be equal to the product of the initial derived effect times the transfer coefficient at the  $j^{\text{th}}$  activity. For instance, the secondary derived effect of research activity  $j$  on activity  $k$ , resulting from an initial amount of research done at research activity  $i$ , is given by  $y_i c_{ij} c_{jk}$ .

The process of derived effects will continue through successive stages until the values of these effects become negligible. This model of the research process expresses the explosive growth features which we have all observed in our own experiences.

We may arrange all the transfer coefficients into a matrix as follows:

$$C = \begin{vmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{vmatrix}$$

where  $c_{ij}$  is the transfer coefficient representing the effect of research activity  $i$  on research activity  $j$ . A matrix such as  $C$  allows for all possible interchanges between research activities. We may further represent the research efforts done on the activities as row vector  $Y$

$$Y = (y_1, y_2, \dots, y_n)$$

where  $y_i$  is the research effort of the  $i^{\text{th}}$  activity. We may now represent the total derived effects by a series of vector matrix products of  $Y$  and  $C$ . Let us denote the vector of total derived effects by a row vector,  $X$ , which is measured in the same units of research effort as  $Y$ . Then we have

$$(1) X = Y(I + C + C^2 + C^3 \dots).$$

The first term on the right hand side of Equation (1), the product of  $Y$  times the identity matrix  $I$ , is just the initial levels of the research activity. The second term,  $YC$ , is the first round of derived effects resulting from the initial level  $Y$ . As described above, the first round of derived effects will have their own derived effects, given by  $YC^2$ , which in turn will have derived effects; and the process continues ad infinitum. We would expect, however, that the coefficients of  $C$  will be such that the successive stages of derived effects steadily diminish. If this is the case, the series of terms on the right will converge to a finite value of  $X$ . The series approaches its limiting value in asymptotic fashion, and the incremental contribution after the first few terms is usually very small.

When the series converges, we may write Equation (1) in the form

$$(2) X = Y [(I-C)]^{-1}$$

and  $(3) Y = X [I-C]$

by making use of a well known algebraic identity.

Equations (2) and (3) are identical in form to those used to study the production process in Leontief Input-Output Analysis. In input-output analysis, one investigates the problem of how to produce a particular final bill of goods, given the interdependencies

that exist between various goods in the process of production. For any net final bill of goods produced for consumption in an economy, a much greater gross bill of goods must be produced in order to have the inputs available to produce the final outputs. The situation is reversed in the research process, where we obtain a greater final output of research effort than our initial input, because of interactions between research activities. The structural similarities between the two situations underlie the mathematical isomorphism between Leontief Input-Output Analysis and our model.

Instead of viewing the research process as a series of stages of derived effects, one can alternatively look at it directly in terms of an input-output structure. The above properties in the transfer coefficients lead to an input-output structure satisfying the following assumptions: the level of research outputs in each field are proportional to the level of inputs in that field, and in addition there may be derived effects between fields which take the form of fractional transfers of research outputs from one field to another. Mathematically these assumptions may be expressed by the formula

$$(4) x_j = k_j y_j + \sum_{i \neq j} x_i c_{ij} \quad i, j = 1, \dots, n$$

where

$x_j$  = total output of the  $j^{\text{th}}$  activity

$y_j$  = total input of the  $j^{\text{th}}$  activity

$k_j$  = the factor of proportionality between inputs and outputs in the  $j^{\text{th}}$  activity

$c_{ij}$  is the transfer coefficient expressing the output of the  $i^{\text{th}}$  activity relevant to the  $j^{\text{th}}$  activity.

Since we have assumed that research output is proportional to research input in a given field, it is convenient to normalize the units of Equation (4) so that a unit of input in a research field leads to a unit of output in that field. Doing this, we get

$$(5) x_j^* = y_j + \sum_{i \neq j} x_i^* c_{ij}^* \quad \text{where } x_i^* = x_i / k_i$$

$$c_{ij}^* = \frac{c_{ij} k_i}{k_j}$$

The vector-matrix formulation of Equation (5) will be identical with that derived above in Equations (2) and (3).

Thus, for an input-output system where there is a strict proportionality between inputs and outputs, the input units may also be used as indices of research outputs. This fact is implicit throughout our analysis at the beginning of this section, in which the above process is viewed as occurring in successive stages. We now consider the final relationships between inputs and outputs.

In the above analysis, we have abstracted from the fact that there must normally be a certain amount of search effort for the derived components of research output in order for them to become effective in any research field. Otherwise they will remain only as potential research outputs for that field. Since this search effort leaves less time available for direct research, the total output corresponding to any given input will be somewhat less than that given by Equation (2).

For various programs of research effort  $Y$ , some emphasizing direct and others, more strongly, pure research, the total derived effects  $X$  can hypothetically be computed and compared by means of Equation (2). In this way, one may seek to evaluate the merits of various balances of effort between pure, applied, and direct research, insofar as they yield total effects on research programs directly related to government activities or national objectives.

The foregoing model is a linear steady state formulation in which the vector  $Y$  and transfer coefficient matrix  $C$  are not obviously measurable for practical purposes. Yet it provides a system suggestive of the underlying structure of research for planning and management purposes. In what terms can the components of  $Y$  be measured and how can the corresponding transfer coefficients be estimated?

One possibility is to define the level of effort of a research activity in terms of man-hours of scientific effort per annum, since ultimately research is a process of applying the trained, imaginative mind of man. Then we may conceive of the transfer coefficients as fractions expressing the man-hours translatable to other research activities and may seek to estimate them in terms of man-hours expended on results which are applicable to other research activities, ignoring the differences between insights of individual researchers as a first approximation or devising some weighting index to deal with the difference in quality between researchers. For this approach it is not clear how expenditures on equipment may be handled. Another possibility is to use dollar value of expenditures on the various research efforts as measures of effort and to estimate the transfer coefficients in terms of dollars expended on results which are applicable to other research activities. Here we have to face the facts that comparable research efforts cost different amounts--depending upon whether they are carried out in universities, government laboratories, or industry; whether they involve the use of heavy equipment; et cetera--and that some price adjustment therefore will be required.

Even after some of these conceptual difficulties are resolved, measurement errors of various kinds are inevitable in any actual empirical application. It is thus important to analyze the effect of these errors on the estimates of the output vector for various input allocations. Detailed analyses of effects of errors in input-output systems are available in the vast literature concerned with Leontief systems.\* For our purposes here it is sufficient to observe that errors in input-output systems have both some welcome and undesirable features. On the one hand, the cumulative effect of the research process implies a cumulative error effect on any estimate, but there is also the tendency for positive and negative errors in different elements of the structural matrix to cancel one another because they affect the final estimate in opposite directions. Furthermore, upper and lower bounds on the size of the error in the vector of research outputs are easily computed for different assumptions on the magnitude of errors in the structural coefficients and various input vectors.

We may illustrate the steady state for our model by means of a simple numerical example. In actual practice we would expect a

\*See especially References (1), (2), and (4).

considerably larger and more intricate system than that which will be discussed here, but this one will serve the purpose of demonstrating the above type of analysis.

Let us suppose that a graph depicting the interrelationships between research activities has been constructed corresponding to that shown in Figure 3. There are seven nodes in the graph

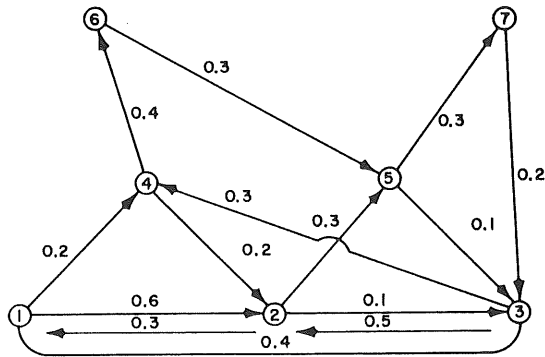


Figure 3. Numerical Example

representing seven research activities arranged to denote a progression on the basic-direct research spectrum. The directed arrows portray the interrelationships among research activities, with the coefficients above each arrow showing the transfer effects of research activities. For example, we see from the graph that a man-hour of work on activity 1 has a derived effect of .6 man-hours of work on activity 2; .4 man-hours on activity 3; and .2 man-hours on activity 4. The matrix of coefficients corresponding to the graph in Figure 3 is given by

$$C^* = \begin{vmatrix} 0 & .6 & .4 & .2 & 0 & 0 & 0 \\ .3 & 0 & .1 & 0 & .3 & 0 & 0 \\ 0 & .5 & 0 & .3 & 0 & 0 & 0 \\ 0 & .2 & 0 & 0 & 0 & .4 & 0 \\ 0 & 0 & .1 & 0 & 0 & 0 & .3 \\ 0 & 0 & 0 & 0 & .3 & 0 & 0 \\ 0 & 0 & .2 & 0 & 0 & 0 & 0 \end{vmatrix}$$

In order to compute the derived effects associated with any initial level of research, we must first calculate  $(I-C^*)^{-1}$ . Subtracting  $C^*$  from the identity matrix and taking the inverse of the resulting product, we get

$$(I-C^*)^{-1} = \begin{vmatrix} 1.40 & 1.32 & 0.76 & 0.51 & 0.46 & 0.20 & 0.14 \\ 0.46 & 1.52 & 0.41 & 0.22 & 0.48 & 0.09 & 0.14 \\ 0.26 & 0.86 & 1.24 & 0.42 & 0.31 & 0.17 & 0.09 \\ 0.10 & 0.32 & 0.11 & 1.05 & 0.22 & 0.42 & 0.06 \\ 0.04 & 0.14 & 0.20 & 0.07 & 1.05 & 0.03 & 0.31 \\ 0.01 & 0.04 & 0.06 & 0.02 & 0.31 & 1.00 & 0.09 \\ 0.05 & 0.17 & 0.25 & 0.08 & 0.06 & 0.03 & 1.01 \end{vmatrix}$$

We may now compute the derived effects for the research situation depicted in Figure 3 by use of Equation (2):

$$X = Y [I-C^*]^{-1}$$

For an initial level of one man-hour for each research activity, i.e.,  $Y = (1, \dots, 1)$ , we have a total effect given by

$$X = [1, \dots, 1] [I-C^*]^{-1} = [2.32, 4.37, 3.03, 2.37, 2.89, 1.94, 1.84].$$

One can see that the total effects from an input of one man-hour to each activity are several times the initial input. Activity 2, which has the largest first stage effects (the sum of the second column of  $C^*$ ) also has the largest total derived effects. It is a central research activity in this structure, having single or multiple interconnections with most of the other activities. Activities 6 and 7, which are the least interconnected, have the smallest total derived effects. Nevertheless, an examination of the matrix  $(I-C^*)^{-1}$  shows that there will be derived effects from each research activity to all other research activities, although some of these effects are negligible. This observation holds true even for a structure of research where most activities have no direct interconnection with one another (only 14 of 49 possible interconnections are non-zero). This

example thus illustrates the necessity of obtaining a system which provides an "overview" of the entire structure of research rather than the consideration of each activity in isolation.

Our model has thus far been developed in a steady state formulation, ignoring the complications arising when time is explicitly considered. Since the development of research is a dynamic process, time ought to be introduced into the model so that the derived effects in one period arise from transfer of research in earlier periods. If we assume for simplicity of analysis that each stage of the derived effects takes one period of time to become effective, then the total effect occurring in period  $t$ ,  $X(t)$ , is given by

$$X(t) = \sum_{j=0}^k C^j Y(t-j),$$

where a finite interval of  $k$  units of time covers all transfers. This is essentially a distributed lag model, with the research occurring in any given period related to that occurring in the immediately preceding period given by the formula

$$X(t) = CX(t-1) + Y(t).$$

In actual practice, we would expect the time for the transfer effects to work themselves out to be different for different research activities. We might expect that the more direct the research activity, the shorter will be the lagged transfers and vice versa. This relation introduces an interesting complication but also increases the burden of the problems of estimating the transfer coefficients.

The roundaboutness of derived effects, both in time and in field of research activity, appears to be characteristic of the research process. For some considerations, the specific time when research results become available may be important; and optimal expansion plans may be sought toward some future goals. Contrariwise, future results may be discounted for current goals.

Throughout the foregoing discussion, we have not tried to make social value judgments concerning the final activities, i.e., concerning the government objectives and the related direct research; and hence we have not suggested optimizations for the present model.

The social worth of end objectives is a valuation problem which must be studied in conjunction with the relevant resource limitations over time, such as budgetary constraints of the support agency and scarcity of specific scientific personnel. If one were to expand the model by introducing explicitly various resources as inputs to the research activities, with constraints upon the total availability of such resources, a preference or utility function could be set up to express the relative social worth of the various final activities. Optimal allocation of research efforts might then be attempted. As a simple example of such a model, we consider the problem of optimally allocating the annual budget  $B$  among various research activities, assuming one has a utility function  $U(y_1, \dots, y_n)$  representing the relative social worth of the final activities. The utility function will in most situations be a *nonlinear* function of the research outputs because of the desire for a balanced research program. We assume as above that each stage of derived effects takes one full period to become effective. Allowance is made for the possible discounting of future returns by the presence of a discount factor  $r$ . Scarcity of research personnel in the different research activities is also taken into account.

Our model may be represented by the programming problem

Maximize

$$U(Y [I+rC + r^2C^2 + \dots]) \\ = U(Y [I+rC]^{-1})$$

subject to

$$\sum_{i=1}^n p_i y_i \leq B$$

$$y_i \leq A_i \quad [i=1, \dots, n]$$

where:

$U$  is the organizational utility function

$C$  is the matrix of transfer coefficients

$Y$  is the vector of research allocation in period  $t$

$r$  is the discount factor  
 $B$  is the total fund available in the budget  
 $P_i$  is an index of the cost of a man-hour of work in research activity  $i$

and

$A_i$  is the maximum amount of scientific man-hours available in a given research activity.

The programming problem may be interpreted as choosing the allocation of research effort which maximizes the value of the discounted derived effects, subject to the constraints that the total budget is not exceeded and that the research effort allocated to any given activity does not exceed the maximum effort available.

This primitive model may be considered to represent the annual budgeting situation. It encompasses mainly short-term considerations in the sense that future budgets and changes in structure are not taken into account. A long-run growth model may be constructed which explicitly considers these complications. Many variations and extensions of the model may be made for optimization purposes. However, before doing so, we should try to implement our basic model and to ascertain empirically whether our underlying model actually describes the real process of research. But we feel assured that the primary aim of this paper has been achieved, namely, to demonstrate that it is possible to construct a model even at the present preliminary state of conceptualization.

#### CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The problems encountered by research supporting agencies in managing large diversified research programs are unquestionably very complex and difficult. Decisions must often be made on what fields of research should be supported and on the level of such support, without any apparent method of assigning comparative value to the research supported. Over the years a great deal of practical experience and intuition in making such decisions has been accumulated by such agencies. These informal procedures and methods are certainly invaluable to successful research management. However,

the tremendous growth in basic knowledge and the increased utilization of this knowledge in the form of new technologies has increased demands for the development of methods and techniques capable of expressing the complex relationships of the research activities to one another and to goals and objectives of the research program.

Of course, any models or other formal devices which are used for the purpose of research management must necessarily be oversimplified. The models that we offer in this paper are not exceptions. Yet the models must not be oversimplified in such a manner as to abstract from the essential nature of the problem studied. In the case of research management, we feel that the logical interdependencies that exist among research activities must be a starting point in the construction of any models.

As a first approach, we offer a graphical model which takes these interdependencies into account by means of directed arcs. The nodes in our graphical model each represent an aggregation of individual research efforts into a single research activity by some rule or principle. Ideally, this rule should satisfy properties that guarantee minimum loss of information resulting from aggregation. The comparison of various alternative rules is a subject for future research. One can use a graph of the type described here with various of its sub-graphs to trace the direct and indirect impacts of various research programs, at least in a qualitative fashion.

A graphical research tree or net is only a first approach in our attempts to characterize the complex interdependencies among research activities. Such a formulation has some obvious advantages over methods which attempt separate comparisons of each research activity with specific government objectives. If we are to deal adequately with these interactions among activities, however, we need a method which is capable of quantitative as well as qualitative analysis. This kind of formulation is given by our matrix models presented in the section on "Presentation of Model in Matrix Form." In order actually to use a model involving quantitative estimates, one must have a method of measuring the outputs of each activity as well as their effects on other activities. Obtaining a good index of research output has been, of course, a central problem in research management. Various attempts to construct such an index have been suggested in the literature, such as counting the number of papers issuing from a research project, weighing experts' opinions,

et cetera, but these have usually been less than satisfactory.

In this paper, we make the simplifying assumptions that research output is proportional to research input measured in man-hours, and that there exists a matrix of numbers between zero and one which expresses the fraction of research spent on any given project applicable to any other project. The latter assumption is equivalent to assuming that the transfer of research between activities is *linear* in nature. In some sense, these assumptions allow us temporarily to bypass the problem of measuring the output of research, but the question of how one measures the transfer effects still needs to be answered. It will be necessary to devise some index which yields an estimate, based on the historical evidence, of the transfer effects and the current feelings of the people conducting the research.

Hopefully, the estimation problems can ultimately be resolved. For the present, we have traced some of the implications of our assumptions concerning the structure of research for research management. The total effect of any given initial allocation of research effort may be expressed by a matrix series of initial and derived effects. Various alternative allocations of research effort emphasizing different mixes of basic and applied research may be investigated, and the resulting derived effects of different allocations may be compared. In addition, the matrix series may be used as a basis for constructing models which incorporate the objectives of the research program and any budgeting or manpower constraints that exist.

After the various problems associated with our basic model have been resolved, our next task would be to relax our assumptions on the linearity of the interactions among research activities and to introduce more complex assumptions involving various nonlinearities. This type of procedure inevitably introduces difficult problems but, we hope, still of mathematical tractability. At any rate, one would hope to obtain from this type of analysis some idea of the biases which result from using the more operational assumption of linearity.

Research management poses many interesting and difficult questions, as stated at the beginning of the paper. It has been essentially an unstructured problem. Hopefully an analysis of models and techniques of the type discussed here will eventually provide policymakers with some useful guides for allocating research funds.

## BIBLIOGRAPHY

1. Evans, W.D., "The Effect of Structural Matrix Errors on Inter-industry Relation Estimates," *Econometrica*, 22 (4) 461 (1954).
2. Hatanaka, M., *The Workability of Input-Output Analysis*, Germany: Ludwigshafen am Rhein (1960).
3. Leontief, W.W. (ed.), *Studies in the Structure of the American Economy*, New York: Oxford University Press (1953).
4. Morgenstern, O. (ed.), *Economic Activity Analysis*, New York: John Wiley & Sons, Inc. (1954).
5. National Science Foundation, Office of Special Studies, *Current Projects on the Economic and Social Implications of Science and Technology*, Annual, Washington (1959-1962).
6. National Science Foundation, Office of Special Studies, *A Selected Bibliography of Research and Development and its Impact on the Economy*, Washington (1958).
7. Price, D., *Government and Science*, New York: Oxford University Press (1962).