MEMORANDUM
RM-6178-PR
APRIL 1970

PPBS, SUBOPTIMIZATION, AND DECENTRALIZATION

Arthur Smithies

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-6178-PR
APRIL 1970

PPBS, SUBOPTIMIZATION, AND DECENTRALIZATION

Arthur Smithies

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of Rand or of the United States Air Force.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

PREFACE

The research reported here is part of a Rand project on military planning and programming systems being conducted for the U.S. Air Force.

This Memorandum has grown out of attempts by the author to assess the effects of various proposals to "control" the defense budget through the imposition of financial constraints by high authority, such as the Secretary of Defense or the Bureau of the Budget acting for the President. Typical of such proposed constraints are budget ceilings on the services, on major programs, or on unified commands.

Such proposals should be appraised in terms of their effect on decisionmaking in a decentralized system whose component parts, particularly the services, have a considerable degree of autonomy. Constraints are needed in any decentralized system. The point of interest concerns the type of constraint that can best contribute to an optimum allocation of defense resources as a whole.

A discussion of this question can take the institutional situation as given; for example, the services can be assumed to retain their existing functions without change. However, the investigator is strongly tempted to consider whether reallocation of functions among services or the strengthening or weakening of central decisionmaking machinery will improve the allocation of defense resources.

Yielding to this temptation has radically altered the form of the present Memorandum from its earlier conception. It now begins with a theoretical discussion of suboptimization and decentralization. With that discussion as a basis, the appropriateness of and the need for various financial controls, constraints, and incentives are then systematically considered.

The author is a Professor of Economics at Harvard University and a consultant to The Rand Corporation. He wishes to thank Rand colleagues Edmund Brunner, Jr., Gene H. Fisher, and Paul R. McClenon for valuable comments and criticisms on an earlier draft.

SUMMARY

This Memorandum first examines the conditions for suboptimization as an aid to decisionmaking in centralized organizations. Optimization by stages is necessary mainly because the objective function of the organization is unknown and has to be discovered by a learning process.

For suboptimization to be possible, marginal rates of substitution among the factors involved must be independent of the value of factors outside the area of suboptimization. When this condition is not fulfilled, as is very often the case, it becomes necessary to devise methods for dealing with the problems of interdependence that remain.

A decentralized system would meet the conditions for suboptimization, but decentralization involves additional problems. It involves the conferring of some degree of autonomy on the lower level decisionmaker. His interest and his point of view are never identical with that of the central authority, or even with some other unit. This conflict of interest produces both benefits and costs for the organization as a whole. In every case, the central authority must devise a system of controls that will hold conflicts between upper and lower levels in the organization within tolerable limits. These controls include incentives, competition, constraints, approvals, vetoes, and indoctrination.

The analysis of suboptimization and decentralization is then applied to various alternative forms of organization in the Department of Defense, and the pros and cons are presented for consideration. These are decentralization by services; full unification; decentralization by service and program; decentralization by theater; strengthening the central controls in the present system; and decentralization by services with some reallocation of functions.

CONTENTS

I. INTRODUCTION

Suboptimization and decentralization are or should be undertaken because of conditions of uncertainty or lack of knowledge. An organization is rarely presented with a clear-cut objective function. Its objectives do not remain constant but grow out of its own experience and out of the changing external environment, including the actions and intentions of friends and enemies.

If everything were known and certain and unchanging, the total problem could be solved without compartmentalization. In fact, in an unchanging world decisionmaking itself would become trivial. In solving practical problems it is necessary to divide the problem into components that are meaningful according to some relevant criteria. Accordingly, foreign policy problems may be considered on a geographical basis. Military problems are often considered according to mode of transportation: air, sea, and land. Education is customarily broken down into student age groups: primary, secondary, and tertiary. In some cases overlapping breakdowns may be needed.

Suboptimization, as the term is used here, involves factoring a total problem, but does not necessarily imply any delegation of authority. Because of his lack of knowledge about objectives, costs, and techniques, a single decisionmaker needs to think of his problem in terms of components, and to build up from partial to total solutions.

Decentralization, on the other hand, involves some degree of delegation of decisionmaking authority. It can arise from the deliberate intent of the central authority, from centrifugal forces within the organization, or from a historical situation such as a federation, where complete centralization is infeasible or undesirable. Decentralization almost inevitably involves some conflict of point of view between central authority and lower decisionmaking levels. The conflict, however, may lead to better results than what occurs in apparently clear-cut centralized processes.

Decentralization and suboptimization are here discussed separately for explanatory reasons. However, they are actually closely related

because the conditions that should be met when suboptimization is undertaken by a single decisionmaker should also govern the design of a decentralized system.

II. SUBOPTIMIZATION

CONDITIONS FOR SUBOPTIMIZATION

The first condition for suboptimization is that the total problem should be factored into components whose outputs are meaningful from the point of view of the whole organization and are, if possible, measurable. These outputs are themselves inputs into a higher level of the decisionmaking hierarchy. In PPBS terms suboptimization is useful if it permits the total problem to be considered in terms of meaningful output-oriented programs. For example, strategic deterrence seems to be a natural candidate for suboptimization since the output of the program is related to a reasonably clear-cut military concept.

The second condition is that there should be no significant interactions between the design of the suboptimized system and other variables in the total system. In technical economic terms, *the necessary and sufficient condition for suboptimization with respect to a group of variables of a total system is that the marginal rates of substitution among members of the group are independent of the value of variables outside the group.*

This condition can be illustrated by some examples. Can education be factored into primary, secondary, and tertiary? For this to be possible, the condition means that an optimum primary system can be designed, with respect to teachers, subjects taught, and buildings and equipment, in a way that is independent of the size or composition of the secondary and tertiary systems. Since virtually all primary school children go on to secondary schools, and since basic skills are needed independently of the composition of the secondary program, it may be reasonable to suppose that the condition is met.

Suboptimizing the secondary system is more problematical. The optimum secondary system will depend on the structure of the primary system. Furthermore, graduates from secondary schools either go to work or go to college. The relative numbers do and should have an effect on the best design of the secondary system; serious mistakes

may be made by ignoring this fact. A possible solution to the dilemma is to suboptimize below the level of secondary education as a whole. For example, preparation of students for college and for the world can be thought of as separate systems with separate outputs.

Another example that suggests itself is to think of "Defense" (D) as depending on the Strategic Capabilities (S) and General Purpose Capability (G). S depends on Minuteman (M) and Polaris (P); G depends on the Army (A) and the surface Navy (N). If prices of M, P, A, and N are known, S and G can be suboptimized provided the rate of substitution between P and M is independent of A and N and *vice versa*. S and G can then be thought of as quantities or can be considered in terms of optimized budgets. The entire defense system can then be considered in terms of S and G. The usefulness of the operation depends on the degree to which S and G can be given clear and distinct military meanings.

Both these examples however, suggest the difficulties of suboptimization as well as its possibilities. In the field of education strong arguments are advanced for the comprehensive high school, despite the logic of having separate ones. In defense, important systems such as carriers and manned bombers have both strategic and general purpose functions.

In fact, it can be safely said that there is hardly any problem where the basic condition is strictly fulfilled. But the need for suboptimization as an aid to decisionmaking remains strong. The problem is to avoid unnecessary difficulties and to discover approximate methods for getting around others.

The foregoing conditions can be stated more precisely with the aid of some symbols. Suppose a decisionmaker wishes to maximize an objective function

$$D = D(x_1\ x_2\ x_3\ x_4)$$

where the x's represent resource inputs.

Suboptimization can be accomplished if it is possible to write,

$$D = F(M,N)$$

where

$$M = M(x_1, x_2)$$

$$N = N(x_3, x_4)$$

or where it is possible to separate the x's in some other way.

This separation of a group variable is possible if and only if the marginal rates of substitution of members of the group are independent of variables outside the group. For example, if the objective function is of the form

$$D = x_1^{\alpha} x_2^{\beta} x_3^{\gamma} x_4^{\zeta} ,$$

$$\frac{\partial x_1}{\partial x_2} = -\frac{\beta}{\alpha} \cdot \frac{x_1}{x_2}$$

which is independent of x_3 and x_4. In this example, then, any group of two or three variables can be separated out. The separation actually undertaken will depend on the significance of the group of variables selected for suboptimization.*

THE TECHNIQUE OF SUBOPTIMIZATION

Let us now examine how a known D function is maximized, through suboptimization and subject to an overall budget constraint.

Let the price of the factors x_1, x_2, ..., be p_1, p_2, ... Without suboptimization, the maximization yields the conditions

$$\frac{\partial D}{\partial x_i} = \lambda p_1 \qquad (1)$$

*For a lucid discussion of this matter see Robert Solow "The Production Function and the Theory of Captial." Review of Economic Studies, 1955, p. 101.

where λ is a Lagrange multiplier. If the D function is factored into M and N as above, this equation can be written

$$\frac{\partial D}{\partial M} \cdot \frac{\partial M}{\partial x_i} = \lambda p_i \qquad (2)$$

or

$$\frac{\partial D}{\partial N} \cdot \frac{\partial N}{\partial x_s} = \lambda p_s$$

depending on whether the subfunction M or N is involved. With suboptimization, the budget is initially allocated between M and N and these functions are maximized separately subject to their own budget constraints. This will lead to a set of equations like

$$\frac{\partial M}{\partial x_i} = \lambda_M p_i \qquad (3)$$

or

$$\frac{\partial N}{\partial x_s} = \lambda_N p_k$$

The budget is then reallocated between M and N until equations (3) correspond with equations (2), and the grand optimum is achieved.*

If all the functions were known, there would be little point in resorting to the cumbersome iterative procedure of suboptimization. The problem would be solved centrally, and, if administration was delegated, explicit and detailed instructions would be given to administrators. This is in fact what Soviet central planning has attempted to do.

A realistic discussion of suboptimization should assume that the form of the D function and even the M and N functions are imperfectly known. However, it will be assumed, at least, initially, that the functions M and N can be factored out, even though their shapes are inadequately known.

---

*Thus far I have borrowed from Alain Enthoven's "Simple Mathematics of Maximization," which appears in Charles J. Hitch and Roland N. McKean, The Economics of Defense in the Nuclear Age, Harvard University Press, Cambridge, 1960.

The first step will be to assign tentative budgets to M and N within the total budget constraint. The second is to establish tentative performance criteria for M and N. These criteria are "outputs" which are produced by the factors of production.

In the military example, the criterion for strategic forces may be the number of enemy targets destroyed. The difficulty of establishing an analogous target for general purpose forces may be the reason why programming has proceeded so slowly in that area. In the education example, the criterion for primary education may be numbers of children achieving prescribed minimum standards in the basic skills.

The tentative nature of these performance criteria cannot be emphasized too strongly. Both at the level of the suboptimizer and at that of the grand optimizer they must be reviewed in the light of experience. At the lowest level, experimentation and research can lead to performance possibilities that have hitherto been unknown -- for example, the possibility of remedial education. At the higher level, decisionmakers are in search of the appropriate "factors of production" for their overall objective. In that connection "numbers of targets destroyed" has been found to be an oversimplistic description of strategic performance.

Such considerations are likely to lead to conclusions that are uncomfortable for the operations researcher; namely, that the "output" of a system or subsystem is likely to involve a number of criteria. In the educational case not only the number but various aspects of the quality of education should be taken into account: how far are able children encouraged and to what extent are retarded children brought up to minimum standards?

In such a case, optimization of a given budget involves first a knowledge of the technological trade-offs, which the suboptimizer is supposed to discover, and also the relative weights that should be allocated to the various criteria. These must be determined by empirical or implicit processes involving the interactions between the suboptimizers and the grand optimizer.

Under these circumstances, M or N must be measured by an index involving the various criteria that should be employed, but such an index is hard to come by. Decisionmakers are likely to have to work with the budgets. At the D level it is necessary to know the implications of an increase in the optimized budget for M or N from the point of view of program effectiveness. In practice, programs are usually advocated and appraised in terms of amount of money spent rather than the physical quantities involved. The optimized budget can in fact be regarded as an index of quantity, provided that both relative prices and the general price level remain substantially unchanged. In times of inflation, such as the present, this problem should be faced by computing the budget in terms of constant prices.

The whole process of optimization should continue to be conducted by an iterative process, with the recognition that uncertainties and lack of knowledge pervade the scene. After an initial budget allocation between M and N, both criteria and techniques are reviewed, and the allocation is revised in the light of that review until, in fact, a satisfactory or perhaps even an optimal situation is achieved.

In a centralized system there are opportunities for some of these operations to be carried out by simulations and staff work. As we shall see, these possibilities become much more limited when there is substantial decentralization.

OBSTACLES TO SUBOPTIMIZATION

Some of the cases where the basic condition for suboptimization are not met -- where marginal rates of substitution among the factors involved are not independent of the value of other factors in the total problem -- can now be usefully examined. Consider the case where a common service, for example, intelligence service, supports both S and G in the defense example. Let this be represented by an additional variable entering into both the M and N functions. So that:

$$M = M(x_1, x_2, z) \text{ and } N = N(x_3, x_4, z)$$

If the total problem is factored in this way the conditions for suboptimization are not fulfilled. However, it may turn out that the marginal rates of substitution between x_1 and x_2 and between x_3 and x_4 are independent of other variables of the system. In that event suboptimization can be carried out with respect to those variables. The determination of z can be regarded as a separate program, whose level is to be determined by the central authority.

Where these conditions do not hold, the level of z will influence the optimum composition of M and N. The optimum composition of z when broken down into components may also be different for M and for N. Such interdependencies as a rule should not lead to abandonment of suboptimization. Recognition of them should lead to organizational arrangements by which they should be taken into account. For example, the committees who suboptimize M and N should include a member with the authority, on behalf of the central management, to determine the level of z -- after discovering the contribution that z can make to M and N.

This argument leads to the conclusion that general support activities should be considered as specific programs. Specialized support, of course, should be included with the programs to which it belongs.

Let us now return to the example of the school system. If M denotes primary education and N secondary, the N system then depends on the subjects taught in M, but M can be suboptimized independently. The functions then are:

$$M(x_1,x_2) \text{ and } N(x_1,x_2,x_3,x_4) \quad .$$

The primary system can be suboptimized but the N system cannot, since we are assuming that the optimal values of x_3 and x_4 depend on x_1 and x_2. However, the problem can be greatly simplified if the operations are performed sequentially. First, optimize M subject to constraints, and find the best relationship between x_1 and x_2. Then solve the total problem, using as variables optimized values of M and x_3 and x_4. Subject to a general budget constraint, the composition of the joint

primary-secondary system can be determined (leaving out the problem of secondary outputs, referred to earlier).

Examples of interdependency can be multiplied. Their numbers and complexity depend on the type of suboptimization attempted. But even if the most efficient method is adopted, interdependencies will remain. They should be taken into account by well-designed committee arrangements, and by some knowledge of the purposes of the organization as a whole on the part of suboptimizers.

III. DECENTRALIZATION

In the discussion thus far we have assumed there is no conflict of interest within the organization under consideration. Decisions may be made by a single central decisionmaker, with suboptimization undertaken purely as a decisionmaking device. Or, if functions are delegated, there is full harmony of interest between the suboptimizer and the grand optimizer.

In practice, delegation implies some degree of what is here called decentralization, involving some conflict of interest between higher and lower levels. At the higher level, it is almost a contradiction to assume that the President and the Secretary of Defense take the same view of the claims of Agriculture on the national budget. It is the responsibility of the President to strike a balance between the rival claims. The Secretaries of Defense and Agriculture are inevitably advocates of the rival claims of their departments.

Other examples are an ambassador who identifies with the country to which he is accredited far more than does the country desk officer, who in turn identifies more than does the Secretary of State. The manager of the Chevrolet Division of General Motors is interested in his own sales and his market shares as well as in his contribution to the profits of the corporation. Generally, the desire for power in a bureaucracy produces some divergence of interest between lower levels and the organization as a whole.

Political federations arise because the component states are unwilling to give up more than a limited part of their sovereignty, and usually not because of a recognition that the national interest is best served by a decentralized process. On the other hand, the planners of an organization may deliberately decide on a decentralized process on the grounds that it does yield the best, if not the most coherent, results. Thus we see that decentralization is not merely an inadvertent consequence of suboptimization. It can result from a historical situation or it can be undertaken deliberately.

Decentralization can be undertaken in varying degrees. At one end of the spectrum it may aim to achieve the same results as suboptimization. At the other it may involve full formal delegation of authority. For example, a new country setting up a new school system might grant full authority to a number of school districts in the hope that a happy blending of community of purpose and of diversity could be achieved.

In terms of symbols, the distinction between suboptimization and decentralization is as follows. The suboptimizer attempts to maximize $M = M(x_1, x_2, \ldots)$, where M is factored out of a total objective function. The decentralized administrator maximizes some function $\bar{M} = \bar{M}(x_1, x_2, \ldots, y_1, y_2, \ldots)$. $\bar{M}$ differs from M for two reasons. First it includes additional variables $y_1, y_2, \ldots$ that are not included in M. For example, with geographical decentralization, the y's could denote local variables that are not taken into account in the national function. Secondly, the weights given to the x's may not be the same in the two cases. For example, a decentralized Air Force may be more inclined to emphasize numbers of aircraft than will the Secretary of Defense.

These divergencies are not necessarily disadvantages. Decentralization usually involves both costs and benefits. The degree of decentralization that is desirable is a matter of judgment arrived at in the light of the particular situation, after weighing the possible benefits and costs of varying degrees of decentralization.

BENEFITS AND COSTS OF DECENTRALIZATION

The first possible benefit of decentralization arises from the fact that the overall objective function is unknown and may remain so. Although this may seem a strange benefit, the diversity that arises from decentralization may be preferable to consistent adherence to an objective function that may turn out to be the wrong one. For this reason, a decentralized school system may be preferable to one that imposes uniform standards. In the former case there will be "good" and "bad" schools but the system on the average may be higher than that of the centralized system. As another example,

decentralization of the armed services served as a useful counterweight to excessive reliance on strategic bombing in the first post-war decade.

Another advantage of decentralization is that individuals have more opportunity and inclination to exercise initiative than they would in a centralized system. Even though an administrator may devote some of his energies to increasing his own influence and prestige, the benefits to the whole organization may be worth that cost. On the other hand, a centralized system gives a greater flexibility and more room for adjustment and experimentation with the system as a whole. A decentralized system can get so ossified that little change is possible.

A further advantage of decentralization is that it reduces the cost of information. In a geographical situation, much local information can be routed direct to operating units and need not go to the central authorities. And even if the central authorities were prepared to pay the cost of the information, the resulting central bureaucracy could be so large as to suffer from diseconomics of scale. Decentralization can preserve loyalties and traditions in the component parts that add to the strength of the whole. On the other hand, in some situations the need to break down tradition may be a valid reason for centralization.

Decentralization can result in duplication of facilities such as R&D, general support, and so forth. But again, some such duplication may be a worthwhile cost of the competition that can often add vitality to the whole organization.

CONTROL OF DECENTRALIZED SYSTEMS

The problem for the central manager is to work out a system whereby the decentralized managers can be encouraged to further the interests of the organization as a whole, and yet one in which they will be restrained from impairing its interest. An overriding condition is an understanding of and a loyalty to the purposes of the organization as a whole. Without such attitudes in key personnel up and down the

hierarchy, no organization is likely to work well; and deliberate processes of training and education may be called for.

Such indoctrination, however thorough, does not guarantee that the decentralized administrator will completely subordinate his interests to those of the organization as a whole. A combination of other devices is needed to achieve as much consistency as possible between his own self-interest and the general interest. Such devices consist of incentives, constraints, competition, directives, and vetoes. The problem is like riding a horse. He is given an incentive to run by the prospect of food or because a companion horse also wants to run; he responds to competition. The reins do not tell him what to do as much as to restrain him from doing something that the rider does not want him to do. But rapport between rider and horse and a willingness to obey are essential to a satisfactory performance. And the horse's own will is never completely submerged by that of the rider's.

Incentives

In a modern corporation the usual and direct way to identify the interests of the lower level manager with those of the corporation is to pay him a bonus based on his contribution to profits. The importance attached to this technique is indicated by the fact that the bonus committee generally consists of the highest officers of the corporation.

This device is not available in government, even in government corporations that make profits. Tradition decrees that salaries must be fixed. Consequently, nonprofit incentives have to be provided to achieve effective performance.

A suggestion has been made that administrators should be allowed to use cost savings to expand their programs. This is not as attractive as it may look at first sight. The budget to which the savings should be added is hard to define. Savings also may be spurious. In the military, a commander might achieve "savings" by reducing fuel inventories, hence dangerously lowering the combat readiness of his force.

Another possible incentive is to permit shifts in bureaucratic power. The constructive empire builders should be permitted to keep at least some of the fruits of victory.

Incentives can also be provided through changing the prices at which lower levels of the hierarchy can buy productive resources. This is the standard method by which the Federal Government influences state and local governments through grants-in-aid. An unconditional grant in effect lowers the price of all resources to the grantee. In such a case the Federal Government is content with the allocation decisions at lower levels, but wants these levels to increase the scale of their operations.

In the United States the most usual grant is conditional. Federal funds are granted to match state funds for some specific purpose, such as the construction of airports or highways, or the control of pollution. In these cases, there is a combination of a general subvention that makes everything cheaper to the states; in addition, however, costs in the grant areas are lowered relative to costs in other areas. The states then have a powerful incentive to expand in the favored areas. The strength of the incentive depends on the matching ratios, which can vary among different programs, and among states according to their capacities to pay.

Within the Federal Government the technique of changing relative prices can also be used or simulated. In defense, military personnel are traditionally supplied "free" to military units. Those units are thus induced to employ more military personnel than they would if they were required to pay for them out of their budgets. From one point of view this system is a source of economic inefficiency. The argument for it is that it maintains a hidden reserve of military manpower that can be (theoretically) drawn on in time of emergency. The principle of free issue is also applied to ammunition and weapons for tactical units, on the grounds that commanders should not be tempted to economize on those items when preparing for or when engaged in combat.

In summary, the device of controlling relative prices takes advantage of the bureaucrat's natural inclination to expand his activities

but it channels that drive in the directions indicated by the interests of the organization as a whole.

Competition

Another way to get the benefits of decentralization is through competition among component units of an organization. It is useful to distinguish between competition among areas where suboptimization occurs and competition within them.

School districts fulfill the condition for suboptimization in the sense that, from the point of view of the quality of education, the optimum school in one district can be designed independently of what is done elsewhere. Competition among districts appears to be a powerful force in improving the quality of education. Communities take legitimate pride in the quality of their systems, compared with neighboring ones. But there are more tangible benefits. Young, vigorous and able families tend to settle into the towns with the best schools. Tax revenues increase and the community prospers.

Things can go wrong, however. Competition may be concentrated not on the quality of education but on more meretricious matters, such as large stadiums, with the consequence that, from the point of view of the country or the region, far more stadium capacity is constructed than is needed. Such competition can arise from the fact that districts are too exposed to each other from a competitive point of view. Each must build a stadium through fear of losing students and population to another district. In that event the process of optimization is interfered with by the creation of new interdependencies. Action by the central authorities to limit stadium construction may be required.

In defense, there can be and is competition between S and G. Within a constrained budget the proponents of each are interested in increasing the contribution that their own program can make to total defense. But they may also be inclined to overemphasize the threats that each type of force is supposed to meet. Moreover, competition

may not stimulate increased efficiency where a competitor believes that the demand for his product is inelastic so that a large part of the benefits of his cost reduction, in terms both of sales and employment, is transferred to his rivals.

Within areas of suboptimization, competition means in effect awarding contracts to the lowest bidder. This corresponds to the practice of design competition in the aerospace industry.

Competition, however, is costly. It means that a number of suborganizations must be supported, with multiplication of R&D and other "overhead facilities" and consequent economic cost. Also, there are conditions under which competition will not work.

Multiplication of R&D activities may, however, have constructive consequences for the improvement of technology. Duplication of production, on the other hand, is much more costly and may have little or no beneficial effect. It may simply result in overproduction of some weapons systems and underemphasis on others, in line with the stadium example noted above.

If the products of the suborganizations are identical, they may prefer to combine and share the market rather than cut each other's throats. This will be especially true if the demand for the whole product is inelastic. Their competitive efforts may then reduce some significant factor such as total employment even though their final product sales may increase. If competition is to be made to work in such a case the rewards for success and penalties for failure would need to be increased. In the defense example, if the competition between P and M meant that one system would be accepted and the other rejected, the competition would be intense. If, on the other hand, it meant incremental increases in one and decreases in the other, there might be a strong inclination on the part of the services to live and let live. Nevertheless, a mix of the two systems may provide the most effective defense.

Further examples can be found in which competition will not work; or where it is not worth its cost. My conclusion therefore is that

competition cannot be considered to be a general prescription for making a decentralized system work.

Constraints

Decentralization, like suboptimization, involves the imposition of budget constraints on lower levels of the hierarchy, and the iterative revisions of these constraints in the light of experience. It is important that the decentralized administrator should be a maximizer, but the imposition of a single general constraint does not guarantee that the right function is maximized.

As pointed out above, there is usually no way of ordering a decisionmaker to maximize some particular function. He must be induced to do so, largely by indirection; incentives and competition help, but particular constraints are also needed. They may be used to impose upper or lower limits on certain strategic variables whose desired magnitude (within limits) can best be determined by the central authority. In the Chevrolet example for instance, a total cost budget constraint might be prescribed. Within that constraint, the manager would be free to vary both quantity produced and unit cost. Such a constraint, however, would do nothing to ensure that the operation was sufficiently profitable. As suggested above, his own interest may induce him to sacrifice profitability for volume. Consequently, central management might decide to place an upper limit on quantity. Under opposite conditions, the Chevrolet manager might need to be restrained from invading the high quality market, in which case a lower limit on production might be needed.

Subject to the fixing of those strategic variables the central management may feel that complete freedom for the Chevrolet manager will be in the interests of the Corporation.

In the military, particular constraints may be needed to prevent commanders from increasing certain features of their programs where such increase may be in their own interests, but not in the interest of defense as a whole. For instance, a particular service may wish

to expand its force level to the detriment of the quality of the force. Consequently ceilings on the numbers of tactical units may be needed.

In other instances the constraint may consist of minimum rather than maximum limits. Such requirements may serve the same purpose as free issue. Services may be required to hold minimum reserves, say of ammunition. These reserves, while necessary in wartime, may do little to increase the peacetime prestige of the service, so that the minimum requirement becomes necessary.

Approvals and Vetoes

In addition to indoctrination, incentives, competition, and constraints there is usually a need for the direct exercise of authority in order to make a decentralized system work.

In connection with grants-in-aid, federal approval is frequently required to assure that federal standards are met. For instance, the national interests demand that airports and highways conform with prescribed safety standards and have uniform signalling systems.

In defense, during the McNamara period at any rate, program change requests and approvals were required when any of the services wished to change segments of programs under their control. This requirement could be justified on the ground that interdependencies among program elements administered by the different services were such that they were not in a position to suboptimize. A further justification would be that service interests diverged from those of the Department of Defense.

IV. POSSIBLE STRUCTURES OF DEFENSE DECISIONMAKING

This discussion of general principles is intended to throw light on various methods of organizing decisionmaking, particularly budgeting in the Department of Defense. In view of this, a number of possible organizational alternatives are now considered, regardless of whether or not they are candidates for serious consideration at the present time.

DECENTRALIZATION BY SERVICES

Service autonomy is the traditional form of defense organization, originally directly under the President and since 1947, under the Secretary of Defense. Despite the changes that have occurred since 1961, decentralization by service remains the central feature of defense organization.

However, that method of organization offends most of the criteria set forth above. In the first place, it violates the conditions for suboptimization. Marginal rates of substitution between factors within given services are not independent of factors administered by other services. For example the Navy operates both strategic weapons and general purpose forces. Substitution between Polaris submarines and carriers should not be independent of the size and composition of the strategic force of the Air Force. Second, although service loyalties and traditions are a source of strength, they can result in large divergencies between the aspirations of particular services and the requirements of the defense program as a whole.

Third, inter-service rivalry can result in duplication of facilities and equipment, especially where rivalries are most intense. The consequence is underemphasis on improvements in those areas where a service is sheltered from encroachments by its rivals. For example, each of the services, if allowed to do so, could overbuild tactical air arms, to the detriment of their area of unique responsibility.

The importance of competition within the military has frequently been underestimated or overlooked. Nevertheless, it appears that the existing services are not the most effective agents of competition. Consider the question of airlift and sealift. The design of this program can be regarded as a legitimate field of suboptimization (although prepositioning should probably be included as a component). Competition of ideas between the proponents of sea and air can work toward the achievement of an optimum mix. The competition may be less effective if it is beclouded by other interests of the Navy and the Air Force respectively.

Most reformers in the past have aimed at reducing service autonomy with perhaps an overconsciousness toward reducing waste and an inadequate appreciation of the usefulness of competition. Now, however, there appears to be a strong trend back toward service autonomy, with a weakening of the controls that have aimed at making the decentralized system work. At the same time there are advocates of a fairly rigid pre-assignment of the budget by service. If the budget were definitely divided among the services at an early stage of the decisionmaking process, this would deprive the defense program as a whole of such benefits as do accrue from inter-service competition.

One proposal by William A. Niskanen* is that, in addition, a new joint service planning group should be set up under the Joint Chiefs of Staff that would exercise much of the authority now reposed in Systems Analysis. In my opinion, such a body could merely transfer the disadvantages of service autonomy to a higher level and work against the "jointness" of the JCS. If we are to have decentralization by services, what is needed at the JCS or the OSD level is a body that will apply the constraints and devise the incentives needed to make the decentralized system work.

---

*"Defense Management After McNamara," Armed Services Journal, February 8, 1969.

FULL UNIFICATION

Full unification would mean that all decisions are determined centrally by the Secretary of Defense or a general staff. Even though the staff is likely to be organized on functional lines, it is supposed to be an extension of the personality of the head of the organization and, in contrast to a decentralized decisionmaker, to act consistently with the objective function of the organization as a whole. The various levels would operate under directives from above. But the quality of the decisions taken would depend critically on the flow of reports and recommendations of the organization.

In this model suboptimization would probably occur at the program level. Members of the general staff would each be assigned responsibility for major programs. For unification in the fullest sense, the separate services would disappear; military operations would be conducted by unified commands. The services could, however, function as supply agencies for particular program elements or they could be replaced by unified supply agencies.

Is full unification desirable according to the criteria worked out above? The main advantage is that it avoids the various conflicts that are produced by decentralization of the services.

An important disadvantage is the cumbersome and possible inefficient nature of a unified general staff. And, as pointed out above, at least some of the benefits of decentralization accrue under the military service system.

A second disadvantage of unification is that the force of competition plays a smaller role in a unified system than it does in a decentralized one. But against this, the wastefulness of competition is reduced.

Perhaps the most important disadvantage of full unification is that it loses the checks and balances of decentralization. It is thus capable of greater success and greater failures. If, as is usually the case, the penalties of failure are greater than the

rewards of success the net effect of unification in this respect should be counted as a disadvantage.

If it were possible to decentralize along functional lines, or to regroup functions among services to conform more closely to the conditions for suboptimization, the case for doing that as against full unification would be very strong.

DECENTRALIZATION BY MAJOR PROGRAMS

Functional decentralization has been proposed by Edmund Brunner of Rand. Under his scheme the present Chiefs of Staff would be replaced by Program Managers who would function in competition with each other for the total defense budget. Programs included would be Strategic, General purpose, Airlift and Sealift, Intelligence, and "Other." Initial budget allocations would be made to each of these program managers, who would then try to maximize the program outputs, subject to their budget constraints. In other words, they would behave as suboptimizers.

They would be assisted in the suboptimization process by competition among the existing services, which would function essentially as supply agencies. For example, the Air Force and the Navy would compete in supplying Minuteman and Polaris missiles for the strategic program. How competition would work with respect to the general purpose mission is less clear, but Brunner relies on competition among the mission commands to achieve an overall optimum. Presumably improvements in effectiveness would lead to budget reallocations that would eventually lead to an overall optimum. There is, however, no guarantee that the process would work out satisfactorily. While there is some competitiveness among programs the dominant relationship is one of complementarity. Consequently savings achieved in one program could result in its own contraction rather than in expansion. This possibility does not encourage efficiency.

Although decentralization by function is a good and perhaps the best basis for decentralization, reliance on competition is inadequate.

Decentralization must be supported by a higher level authority that provides incentives, constraints, and encouragement in line with the previous argument.

Furthermore, competition among the services in the various program areas may not be intense. The standard examples of Minuteman versus Polaris and airlift versus sealift are the ones that come immediately to mind. Others come less readily.

In the absence of testing, experimental or actual, it is impossible to say how Brunner's proposal would work. In the meantime, it can be questioned on the grounds that it does not conform well to our a priori criteria.

DECENTRALIZATION BY SERVICES AND PROGRAMS

Another proposal, made by Herbert Rosensweig while he was in the Office of the Secretary of Defense, amounts to a combination of decentralization by services and by programs.

First, it is proposed that the traditional constraints on the services through the method of prescribing force levels should be replaced by financial ceilings. This change is based on the view that the services tend to goldplate their equipment if they are constrained on numbers. On the other hand, we have seen that competition can lead to increases in numbers at the expense of quality. Both kinds of constraint would probably be needed. But whatever is done, decentralization by services is open to the objections raised above.

The Rosensweig proposal prescribes budget ceilings not only on the services as a whole, but on programs within each service. All of this amounts to prior determination of the budgets in each service, each program, and each cell of the program-service matrix.

It is hard to see how such an elaborate web of ceilings could lend itself to the process of iterative adjustment that is needed in a decentralized system. The only way in which I can conceive of the system working is for all the allocating to be done initially and definitely by the OSD staff. The only discretion left to the services

would be to seek improvements within the cells of the matrix. For example, the Air Force could seek improvements in Tactical Air and the Army in its ground forces. But they would have nothing to say about the interactions of these forces with other components of the defense program as a whole.

If this is the interpretation to be placed on the proposal, it comes close to full unification and should be considered on those terms.

DECENTRALIZATION BY THEATER

A further possibility is that the unified commands in the various theaters should provide the basis of decentralization. The unified commands meet the condition for suboptimization better than do the individual services but not necessarily better than do major programs. Competition for resources among theater commanders may achieve a good balance from the point of view of the attainment of national objectives.

Of course decentralization would relate mainly to Operation and Maintenance. R&D and Procurement would remain centralized, or decentralized on a functional rather than a theater basis. Such an arrangement would resemble the organization of the Navy, with theaters corresponding to fleets, and the centralized functions corresponding to bureaus.

From the point of view of meeting the conditions of suboptimization, decentralization by theater seems as valid a basis as decentralization by mission. This is but one instance of the pervasive regional versus functional question. It may be possible to employ both methods side by side, with crosswalks between them at various levels of the hierarchy. Alternatively, the regional problem may lend itself better to handling by the central staff than the functional. Then decentralization would be undertaken on a functional basis, but the process would be subjected to constraints or directives that would safeguard regional interests. Perhaps that is the way the theater

question should be handled. In fact it is at present handled mainly by consultation between theater commanders and the JCS. Whether the process is adequate is a subject that deserves further study.

There is, however, a major obstacle to organizing the whole defense program on the basis of unified commands, especially the theater commands. Such a system is not equipped to deal with mobile forces located in the United States that can supply any theater as the need arises, through airlift or sealift. Nor is it likely to take due account of the possibilities of transferring forces from one theater to another. Forces located throughout the world should be placed in competition with mobile forces in the United States. Decentralization by mission can embrace such questions, but autonomy for unified commands can result in waste through paying too little attention to the possible economies of mobility.

Nevertheless, it is important to impose budgetary responsibility on the unified commanders after the areas of their responsibility have been established. This would be done as a supplement to a functional or service basis and would relate primarily to Operation and Maintenance.

STRENGTHENING OSD AUTHORITY

The intent of the McNamara reforms was to superimpose programming under the Secretary of Defense on the traditional system, as the dominant element in the decisionmaking process. The program budget was to be translated into the traditional categories through "crosswalks." The latter were to be retained for administrative convenience within the departments and because of Congressional interest in the existing system.

Unfortunately for the programming system, its dominance was never fully established. While the budgets of the services were supposed to be translations into budget categories of approved programs (and program changes), this intent was never fully realized.

The services in fact retained the ability to add to their budget requests through the regular process. Consequently, the budgets of the McNamara period resulted from the old system and the new working in awkward double harness. The programming system, instead of dominating the whole process, may well have had its main impact in influencing the design and procurement of particular weapons systems.*

Whether the decisions reached in these complex processes were good or bad can be debated at length and inconclusively. What can be said, however, is that the present arrangement is unstable from the point of view of the balance-of-power within the DoD. During the past year, that instability has been evidenced by a pronounced strengthening of the services in the budgeting process. If, in the future, the program influence is to be more vigorously asserted, it will be necessary to alter procedures so as to increase the influence and strength of central staff agencies, such as Systems Analysis and DDR&E in relation to the services.

Some changes in present procedures that could contribute to this objective are:

(1) The OSD staff might be strengthened along program lines. There should be staff directors for major programs and subprograms.

(2) The Joint Staff of the JCS should be selected with particular stress on experience in unified commands and other experience that would tend to encourage officers to take a joint rather than a particular service point of view.

(3) The defense program and budget would result from a process of suboptimization by program rather than decentralization by service. Initial budget constraints would be placed on programs and would be revised in the light of experience and analysis.

---

*This paragraph, while conforming to my own impression, relies heavily on a careful study by John P. Crecine, Defense Budgeting: Organizational Adaptations to External Constraints, The Rand Corporation, RM-6121-PR, March 1970.

(4) The services under these arrangements would function more as supply agencies and less as strategy-forming agencies than they do at present. They would receive budget allocations based on program decisions. But their effectiveness and competitiveness as supply agencies would affect the allocations they receive.

(5) Appropriations should be made to the Secretary of Defense rather than to the services. He should allocate funds among the services, after the budget is enacted and should have freedom to transfer among services after allocations have been made.

(6) Strenuous efforts should be made to persuade the Congress to accept an appropriation structure that corresponds with the program structure.

These suggestions move in the direction of the Brunner proposal. The services, as supply agencies, would perform much the same role as they do in his scheme. On the programming side, however, my suggestion is less radical than Brunner's. Programming remains a staff operation in the OSD. In contrast, Brunner proposes a decentralized programming structure, with program directors having more autonomy than is here envisaged. It is far from clear, even to its author, how his process would work. It could increase present conflicts and tensions between programmers and services, without constructive results. The present suggestion at least builds directly on experience. It retains the benefits of what has gone before, but does not resolve all the difficulties.

DECENTRALIZATION BY SERVICES WITH REALLOCATION OF SERVICE FUNCTIONS

All of the suggestions so far discussed involve conflict between the OSD and the services. As I have maintained, conflict is inherent in any decentralized organization, but with the existing division of functions among services it is excessive and unconstructive. Moreover, the present division of functions results partly in the type of interservice competition that is wasteful rather than constructive. Some of the foregoing proposals, and also the McNamara system, tend

to degrade the services into supply agencies. There is some question whether this is the best use of institutions that are military in their training, aptitude, and traditions. The best use of military personnel may well be to perform missions and to devise strategies.

Consequently, I suggest for consideration an arrangement that would rely on decentralization by services, which in the light of service traditions seems to be the natural division. However, there would be some redistribution of functions among the services so that they would more closely conform to the suboptimization conditions.

The supply function would be performed by supply agencies whose orientation was towards business management. They would be military or civilian as the circumstances dictated. Examples of such agencies have been increasing in recent years. Military personnel is one of the oldest examples of the separation of a supply function. More recently the Defense Supply Agency has been set up to supply common use items to all the services. A separate agency has been established to procure petroleum, oil, and lubricants and sell them to the services.

A possible regrouping of forces would be:

Strategic Forces - consisting of SAC, Polaris, Air Defense, and ABM.

Tactical Land Forces - consisting of the Army and Tactical Air.

Tactical Sea Forces - consisting of the Surface Navy and A.S.W.

Airlift and Sealift

Intelligence

Possible supply agencies would include:

Military Personnel

R&D

Aerospace Procurement

Naval Procurement

Defense Supply Agency

POL

Construction

Under such an arrangement the Joint Chiefs of Staff would consist of the Chiefs of Staff of the forces (but not the heads of supply agencies). Since the activities of the regrouped services would be largely complementary rather than competitive, they could be expected to take a more unified point of view than the present JCS, and conflicts between them and the OSD Staff could be expected to diminish.

Of course, no regrouping could deal with all problems of interdependency. Both the carrier and the manned bomber serve dual strategic and tactical purposes and there is nothing to do except recognize that fact.

On the supply side, a unified military personnel agency may become important to the extent that volunteer armed forces with close contact with the labor market are to be relied on. A unified manpower procurement agency, however, need not be inconsistent with the preservation of service traditions. A unified R&D agency would presumably be concerned with nonmission-oriented functions. Mission-oriented R&D would remain under the control of the military services.

Under this arrangement both initial and final budget allocations would be made to the reconstituted services (with freedom to transfer remaining with the Secretary of Defense). The supplying agencies would receive most of their money as payment for goods and services supplied to the services. Direct appropriations to them would cover fixed investment and working capital.

This Memorandum has attempted to present alternative decision-making arrangements, rather than to make definite recommendations. However, it can be said that this last possibility seems to meet the requirements of a decentralized system better than any of the others.

DOCUMENT CONTROL DATA

1. ORIGINATING ACTIVITY The Rand Corporation	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE PPBS, SUBOPTIMIZATION, AND DECENTRALIZATION	
4. AUTHOR(S) (Last name, first name, initial) Smithies, Arthur	

5. REPORT DATE April 1970	6a. TOTAL NO. OF PAGES 37	6b. NO. OF REFS. ---
7. CONTRACT OR GRANT NO. F44620-67-C-0045	8. ORIGINATOR'S REPORT NO. RM-6178-PR	

9a. AVAILABILITY/LIMITATION NOTICES DDC-1	9b. SPONSORING AGENCY United States Air Force Project RAND

10. ABSTRACT

A theoretical discussion of suboptimization and decentralization to consider the appropriateness of various financial controls, constraints, and incentives. For suboptimization to be useful as an aid to decisionmaking in centralized organizations, marginal rates of substitution among the factors involved must be independent of the value of factors outside the area of suboptimization. A decentralized system would appear to meet these conditions, but there are additional problems--in every case the central authority must devise a system of controls that will hold conflicts between upper and lower levels in the organization within tolerable limits. The analysis is applied to various alternative forms of organization in the Department of Defense: decentralization by services, by service and program, and by services with some reallocation of functions; decentralization by theater; full unification; and decentralization by strengthening the central controls of the present system.

11. KEY WORDS

PPBS (Planning, Programming, Budgeting, Systems)
Organization Theory
Defense Department
Decisionmaking
Military Planning