Volume Title: Issues in Defense Economics

Volume Author/Editor: Roland N. McKean, editor

Volume Publisher: NBER

Volume ISBN: 0-87014-490-1

Volume URL: http://www.nber.org/books/mcke67-1

Publication Date: 1967

Chapter Title: Collective Goods, Comparative Advantage, and Alliance Efficiency

Chapter Author: Mancur Olson, Jr., , Richard Zeckhauser

Chapter URL: http://www.nber.org/chapters/c5159

Chapter pages in book: (p. 25 - 64)

Collective Goods, Comparative Advantage, and Alliance Efficiency

MANCUR OLSON, JR.
AND
RICHARD ZECKHAUSER

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
HARVARD UNIVERSITY

I

Defense is a classic example of a public or collective good. Our understanding of the implications of this fact depends in considerable part upon the state of the theory of collective goods. In recent years a number of leading economic theorists have turned their attention to this theory and they have made some important advances. Still, by comparison with the theory of private goods, the theory of public goods has been sadly neglected. There are a number of respects in which this theory as usually understood is ambiguous and mistaken, and these theoretical difficulties must be resolved before the implications of the fact that defense is a collective good can be adequately understood.

Any attempt to clarify or improve the theory of public goods should in turn begin with a clear definition of the concept of a public good. As John Head has shown,[1] two distinct defining characteristics have been implicit in most discussions of public goods. One defining characteristic of a public good is that it is not possible, or at any rate

NOTE: This article was written before the first named author was employed by the government and does not, of course, necessarily reflect official views or policy. The authors are especially thankful to William J. Baumol, James M. Buchanan, John Hauser, Wallace E. Oates, and Robert T. Smith for helpful criticisms, but all of these critics disagreed with parts of our argument and share no responsibility for its shortcomings.

[1] "Public Goods and Public Policy," *Public Finance*, No. 3, 1962, pp. 197–219.

economically feasible, to exclude nonpurchasers from the consumption of it. If, for example, the East Coast is protected against amphibious invasion, Midwesterners cannot feasibly be excluded from the benefits of this good. The other defining characteristic of a public good is that if provided to one individual in a group, it can be provided to the others at little or no marginal cost. To put it differently, it is a good of a type that can be consumed by an additional individual without any significant reduction in the amount available to the original consumers. Defense usually has this characteristic. So would a bridge in a sparsely traveled region where even the smallest technically feasible bridge would entail excess capacity. As the possibility of a toll barrier excluding nonpurchasers at such a bridge reveals, the two defining characteristics of a public good need not always be present together. A good that has only one of the aforementioned characteristics can usefully be analyzed as a collective good, though these two characteristics are most commonly found together.

It is immediately evident that the concept of external economies is included in this definition of a collective good. An external economy has long been defined as a benefit which nonpurchasers cannot be kept from enjoying and which therefore cannot fetch a price in the market. By virtue of this infeasibility of exclusion it is also a collective good. Many types of external economies are also such that when others come to enjoy them there may be no proportionate reduction in the amount available to those who enjoyed them from the beginning. This is the case with immunization against contagious diseases, for example, where the immunity from disease acquired by part of the population may benefit others in the population without reducing the benefits to those who benefited originally.[2]

A consideration of external economies and the defining characteristics of public goods suggests that "public goods" might better be called "collective goods." The phrase "public good" suggests the public sector, but such goods are also provided by individuals and private groups and organizations. Obviously, an individual firm or consumer can provide external economies. Television, one of the better examples of public goods, is often provided by private firms. Similarly, as one of

[2] An external diseconomy can similarly be regarded as a collective or public "bad," and its prevention or minimization as a collective good.

the present writers has argued elsewhere,[3] almost any organization provides some collective goods. Labor unions provide many collective goods to their members and pressure group lobbies provide collective goods to the groups for whom they win favors. Indeed, a collective or public good exists whenever any group or organization strives to obtain a purpose or objective such that, if it is achieved for any one member, it is automatically available, or can be made available without significant cost, to the other members of the group or organization. Surely it is better to classify all goods with similar analytical characteristics together, rather than to distinguish on institutional grounds those provided in a given case by governments. Thus we shall here use the phrase "collective good" rather than its more traditional counterpart, "public good."

The fact that collective goods are inherent in organizational or group efforts to attain a common objective has a special importance in the international context. For it suggests that whenever two or more nations have a common enemy, and one of the two engages, weakens, or destroys the common enemy, there will be a collective good or external economy the other nation may also enjoy. And as the ancient maxim that "mine enemy's enemy is my friend" suggests, this could be true even in the absence of an alliance treaty, a common ideology, or even cordial relationships. We therefore conclude that international organizations, international cooperation, and military alliances, whether tacit or formal, produce collective goods.

If military alliances and other forms of international collaboration for common purposes produce collective goods, what are the implications of this fact? The theory of collective goods and externalities should make these implications clear, but unhappily recent writings on collective goods and externalities seem to differ on some of the basic propositions of the theory. One of the most interesting disagreements concerns the comparison between the quantity of a collective good that is supplied under independent adjustment in a "market," and the Pareto-optimal quantity of that good. The present writers have argued elsewhere that nations in an alliance tend to provide only suboptimal amounts of the collective goods they seek,[4] and the logic of Pigou's classic model of ex-

[3] Mancur Olson, Jr., *The Logic of Collective Action: Public Goods and the Theory of Groups,* Cambridge, Mass., 1965.

[4] Mancur Olson, Jr., and Richard Zeckhauser, "An Economic Theory of Alliances," *Review of Economics and Statistics,* August 1966, pp. 266–279.

ternal economies would also seem to imply that too little of the alliance good would be provided. This conclusion, however, at first sight appears to be inconsistent with some important recent papers by James Buchanan and Milton Kafoglis and by William Baumol.[5]

II

Let us first consider the important contribution by Buchanan and Kafoglis. They argued in a seminal paper in the *American Economic Review* that it was *not* necessarily true that independent adjustment or market organization of an activity with substantial external economies would fail to provide an adequate outlay of resources on that activity. In their view, the independent behavior of the individual parties with externalities or collective goods might lead to an adequate or even a supraoptimal level of resources devoted to an activity with external economies. They claimed that this meant there was a flaw in the "orthodox" argument for the collectivization of activities with substantial external economies. As they put it:

> The theory of economic policy upon which arguments for the collectivization of any activity must be based embodies the prediction that the behavior of individuals in markets does not produce socially desirable results. In the orthodox analysis, this prediction stems from the presence of significant externalities that the market is presumed unable to internalize. In his independent behavior, the individual is assumed to take into account only the effects of his actions on his own utility or that of his family group. From this it follows that, if private behavior exerts Pareto-relevant external economies, the market-generated supply of resources to the activity in question falls short of the "social optimum," as this is defined by Pareto criteria. . . .
>
> In this note we shall demonstrate that this orthodox policy implication is not completely general, and that, in certain circumstances, it may be in substantial error. Independent or market organization of an activity that is acknowledged to embody relevant external economies need not result in an undersupply of aggregate resource inputs, relative to that amount required to satisfy the necessary marginal conditions for Pareto-optimality.[6]

[5] J. M. Buchanan and M. Z. Kafoglis, "A Note on Public Goods Supply," *American Economic Review,* June 1963, pp. 403–414; W. J. Baumol, "External Economies and Second Order Optimality Conditions," *American Economic Review,* June 1964, pp. 358–372.

[6] Buchanan and Kafoglis, "Public Goods Supply," p. 403. Buchanan and Kafoglis were not at all explicit about what they believed the policy implications

Buchanan and Kafoglis point out that their interest in this problem was stimulated by an attempt to "explain" the fact that expenditures on medical services have expanded more rapidly in the United States than in Britain in the period since British medical care has been collectivized. In their opinion, traditional Pigovian externality analysis would have suggested the reverse. Their own model, they find, is at least consistent with the relatively greater increase in expenditures on medical care under the market-organized system in the United States. They also speculate that expenditures on private firearms and private policemen might increase if public provision of police services were to cease, and this again would appear to contradict Pigou's analysis and support their own. Finally, though Buchanan and Kafoglis did not discuss defense or alliances, it would follow if their analysis is correct that there is no necessary tendency for allied nations to provide too little alliance defense.

Buchanan and Kafoglis consider two separate types of cases of externalities or collective goods. If the consumption of a good or service by one person, say person B, exerts externalities on another person or persons, say person A, but A's consumption of this good or service exerts no externalities on B, then the relationship is said to be "nonreciprocal." In the "reciprocal" case by contrast the activity of each party provides externalities to the others. Buchanan and Kafoglis argue by example and start with a case in which any immunization shots obtained by individual B not only protect his health but also provide a spillover, in the form of reduced chances of being subject to contagion, to individual A. But this relationship is assumed to be nonreciprocal, so immunization shots given A benefit only him. They then consider cases in which an immunization shot for B is a perfect substitute for a shot for A, a better than perfect substitute for such a shot, or a less than perfect substitute, and find that in some situations independent adjustment will lead to the same total number of shots as an ideal collective arrangement, whereas in some other situations it will lead to a smaller number of shots, and in others to larger numbers of shots.

of their argument were, and there is the possibility that the reader may read policy positions into their argument that they never intended. On this problem see Section IV below, including notes 16 and 21. There is a further statement of at least Buchanan's views on this topic and its policy implications in James M. Buchanan and Gordon Tullock, "Public and Private Interaction Under Reciprocal Externality," in *The Public Economy of Urban Communities,* edited by Julius Margolis, Washington and Baltimore, 1965, pp. 52–73. We had not read the aforementioned article when our paper was first presented.

The simplest and most instructive case is that in which an immunization shot for B is a perfect substitute for such a shot for A, so we shall consider Buchanan and Kafoglis' geometric treatment of such a case here. Buchanan and Kafoglis neglect income effects in order that they may use marginal evaluation curves as demand curves. The valuation individual *A* places upon protection from contagious disease at the margin is therefore given by demand curve D_A in Figure 1-A, and the same is true of *B*'s demand curve D_B in Figure 1-B. The marginal cost of shots is assumed constant and the same for both and is represented in each figure by the line *CC*. If *A* initially considered himself in isolation he would purchase *OQ* of disease protection. *B*, who in this non-reciprocal case by assumption receives no spillover from *A*, purchases quantity *OM* of disease protection. When *A* becomes aware of *B* he

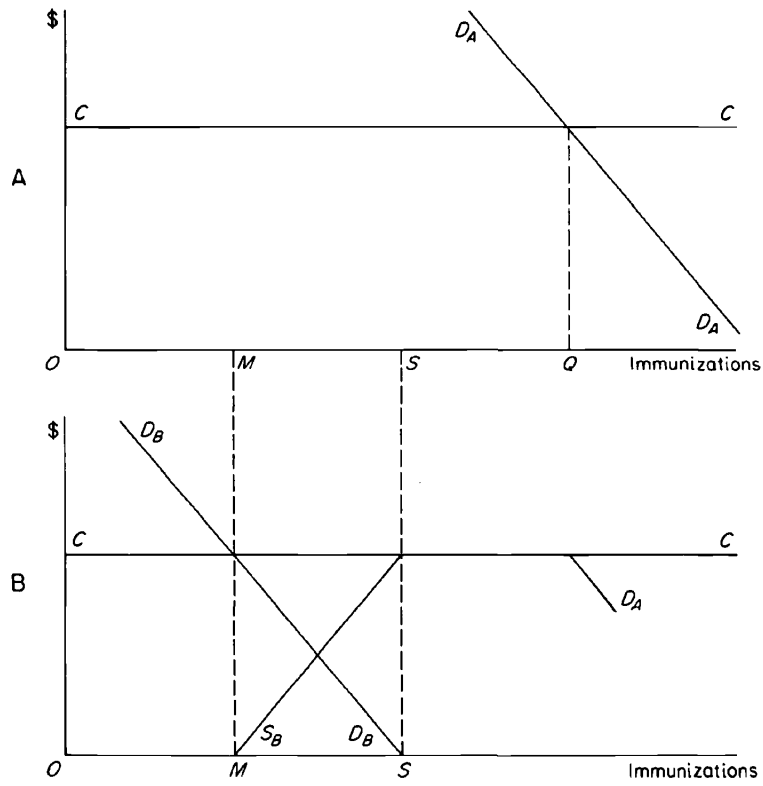


Figure 1

will reduce his purchase of immunization by one unit for every unit that B has obtained, i.e., he will reduce his purchase by an amount equal to OM, and purchase only MQ himself. This total outlay, $OM + MQ$, or OQ, is the total equilibrium outlay with independent adjustment.

Buchanan and Kafoglis then compare this market equilibrium with the Pareto-optimal level of provision. Since each shot given to B provides immunization both for himself and for A, both A and B can be made better off by taxing A and using the proceeds to induce B to buy additional immunization shots. The value to A of immunizations of B is given by the marginal cost curve CC until it touches A's demand curve, at which point that demand curve indicates the value of additional immunizations for him. B's supply curve of immunization protection for A will be given by the line S_B, which is obtained by subtracting his demand curve from the marginal cost curve. The ideal collective outlay is achieved where B's supply curve intersects A's demand curve, that is when B obtains OS of immunization and A purchases an additional SQ of disease protection, for a total of OQ.

What happens in this particular example is that under the ideal collective arrangement A's intake of immunization shots is reduced by one for every extra shot B is induced to take, with the result that exactly the same total number of shots is bought under the Pareto-optimal collective arrangement as is bought under the independent market adjustment. Buchanan and Kafoglis therefore draw the general conclusion that independent market adjustment may sometimes provide as large an outlay on a good with external economies as an ideal collective arrangement. By varying the conditions slightly, they also show that independent adjustment can sometimes bring about a larger and sometimes a smaller outlay on a collective good than an ideal collective arrangement would employ. Suppose an immunization shot for B were a more than perfect substitute for a shot for A. Then independent adjustment would lead to a larger outlay on immunization shots than an ideal collective arrangement. This is the case because each extra shot B takes under the ideal collective arrangement makes it appropriate to reduce A's outlay by more than one shot.

When Buchanan and Kafoglis develop examples dealing with the reciprocal case, where a shot for A helps B as well as *vice versa,* they use a hypothetical payoff matrix, showing the payoffs to each of the two individuals as one or the other, or both, get additional immunization

shots. Still assuming nonstrategic behavior, they show that in their example that the action of the two individuals can result in a larger expenditure on shots than an ideal or Pareto-optimal arrangement would. They also show with their example that the final output, i.e., the amount of immunization or disease protection, that is obtained under independent adjustment can be larger than the amount that would be obtained under a Pareto-optimal arrangement. They argue that this "suggests that, under independent adjustment, the consumption of final output is overextended." [7]

III

We shall later contend that Buchanan and Kafoglis' very challenging argument is not satisfactory in every respect, but it will be necessary first to turn to an interesting and very important paper by William Baumol, which appeared in the form of a communication in the *American Economic Review* about the Buchanan and Kafoglis contribution. Professor Baumol offered "an explanation of the Buchanan-Kafoglis result" [8] which would "help to substantiate the validity and significance of [their] argument." [9] One part of Baumol's contribution consisted of the discovery and demonstration of the important point that, when externalities are present, the second-order conditions are often violated, and there are often multiple maxima and corner solutions.[10]

The other part of Baumol's contribution is the more relevant here, for it seems to offer a general explanation of the Buchanan-Kafoglis examples. Baumol showed that on a social welfare function on which there are no multiple maxima, and where the second-order conditions

[7] Buchanan and Kafoglis, "Public Goods Supply," p. 411.

[8] Baumol, "External Economies," p. 360.

[9] *Ibid.*, p. 359.

[10] The plausibility of this contention is easily evident in a situation where, in the absence of externalities, there would be decreasing returns to scale to each of two activities. If these activities provide external economies to each other, however, and the externality of each activity increases with its output, this may after some point bring about increasing returns for the two activities together. This could mean a local maximum at low levels of output and another maximum at higher or conceivably even infinite levels of output, with a minimum point in between. Similarly, if two activities exert external *dis*economies upon each other, it may be better to have *either* of the activities operate at a high level and the other activity at a low (or perhaps zero) level, than it would to have both operate at something like the same level. In such a situation there would also be two (or more) maxima and a minimum point in between.

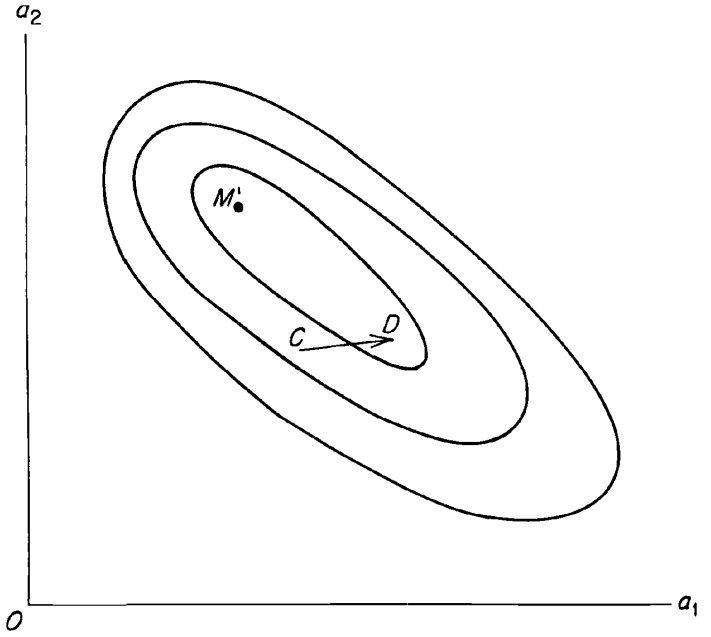


Figure 2

for a maximum are fulfilled, the marginal rules for maximization can lead the society away from the peak of the social welfare function. Consider Figure 2, in which the level of welfare achieved is a function of the levels of activities a_1 and a_2. The utility surface has a unique maximum at M'. A movement from C to D would be consistent with the marginal conditions for maximization, for D is at a higher level of welfare than C. But while D is higher than C, it is nonetheless farther from the highest point! Going up the hill can move one away from the top: an improvement can leave one farther from the ideal than before.[11]

[11] We are thankful to Richard N. Cooper for pointing out in his comment on our paper that the hill in the figure should be drawn in such a way that its length is not parallel to either of the axes (as it had been in our first draft), but should rather run diagonally to the axes as it now does. Even if the length of the hill is parallel to an axis it is possible to move up the hill while moving farther from the peak, and thus demonstrate the argument of the text. But in such cases there would be negative returns at the margin to one of the activities. If on the other hand the length of the hill runs diagonally, it can be possible to move away from the peak of the hill without negative marginal returns to either activity. It is also possible, if some less regularly shaped hills are drawn, to

The spirit of Professor Baumol's paper is quite different from that of Buchanan-Kafoglis, since Baumol did not appear to be concerned as much with whether and when collectivization might be appropriate as with the dangers of piecemeal planning procedures and the inadequacy of simple Pigovian analysis. Still, his argument seemed to add strength to Buchanan and Kafoglis' contribution, for it showed that the rule that an activity should be expanded when marginal social benefit exceeded marginal social cost could lead a society away from the welfare ideal. Thus Baumol argued that his analysis provided "an explanation for the Buchanan-Kafoglis result," [12] and concluded, "Buchanan and Kafoglis have raised a fundamental point in questioning whether external economies need always call for increased outputs." [13]

Baumol's rigorous, general, and mainly mathematical analysis therefore seemed to explain and substantiate the points suggested by Buchanan and Kafoglis' practical and hypothetical examples. Yet in fact these challenging contributions, important as they are, nevertheless can be shown to be ambiguous, misleading, and incomplete in several important respects. Moreover, they both contain contributions of considerable importance which the reader, with some difficulty, has to discover for himself.

IV

The first thing that needs to be said is that Baumol's discovery of the improvement that takes one farther away from the ideal is much more general and significant than he points out. It can occur whether there are externalities or not. Whenever a lack of knowledge, a change in technology, taxation, monopoly, or anything else has left a society off the ideal point on its social welfare function, it can easily be the case that the marginal rules can lead it astray—that is, to improvements that leave it farther from the ideal point.[14] It is of course true that continued

show that it is possible to rise along the path of steepest ascent of the hill, yet move away from the peak, even when the hill has no local maxima. We are also thankful to Michael Intriligator and William Vickrey for stimulating comments on this point.

[12] Baumol, "External Economies," p. 360.

[13] *Ibid.*, p. 367.

[14] It may be the case that knowledge is particularly scarce where externalities are concerned. To the extent that this is the case, Baumol's argument has a particular importance when externalities are present.

upward movements or improvements will ultimately lead to at least a local maximum. But if, as is usually the case, resource reallocations are costly, and if, as is all too common, vested interests that resist change tend to build up around any given allocation of resources, a long series of circuitous movements toward the ideal may well be undesirable, if not practically impossible.[15] So improvements that lead a society away from its ideal may in a dynamic situation actually reduce its long-run welfare. In other words, if the third best is nearer to the ideal than the second best, it may ultimately prove costly to move from the third to the second best. Baumol's point seems to us to be almost as general or far reaching in its relevance as the well-known theory of the second best.

But general as it may be, it neither explains nor substantiates the Buchanan-Kafoglis examples, nor does it reveal that these examples should suggest conclusions quite different from those that appear to be suggested in the Buchanan-Kafoglis contribution. This can be seen most easily by thinking back to the Buchanan-Kafoglis nonreciprocal case, where an immunization shot for B was a substitute for a shot for A, but not *vice versa,* and where the examples suggested that the expenditure on immunization shots could be as great as or even greater under independent market adjustment than under Pareto-optimal collective arrangements. The problem here is to understand the relevance of the assumption that an immunization shot for B gives immunity to A though the reverse is not true. This assumption means—and this is the simple essence of the problem—that there are two different production functions or productive processes that can be used to produce immunity from disease in the example, and that these two production processes vary in their efficiency. If a shot for B protects both B and A and a shot for A protects only A, the presumption is that B is a more efficient producer of immunity than A. If, as Buchanan and Kafoglis sometimes assume, a shot for B is a perfect or more than perfect substitute for a shot for A, and B also places a positive value on immunity, it is *necessarily* the case that B is a more efficient producer of immunity than A. All this can be easily seen by returning to Figure 1, where it is obvious that there is a larger output of immunity under the ideal collective ar-

[15] We are thankful to Gardner Patterson for reminding us of the importance of such obstacles to resource reallocations in this context.

rangements ($OQ + OS$) than under independent market adjustment ($OQ + OM$), even though the total number of shots taken is the same in both cases. The reason is that some shots have been reallocated from A to B. Indeed, Buchanan and Kafoglis point this out themselves.

If there are two different ways of producing a product, such as immunity from disease, and one is more efficient than the other, it will often be the case that more will be spent on the good if the less efficient method of production is employed, even though less is produced. If the users of a good have a demand curve for the good with an elasticity of less than unity in the relevant range, then increasing the cost of this good to its users will increase their expenditure on the good. The use of a less efficient process of producing a good raises its supply curve and makes it more costly for its users (consumers). This then is the explanation of all of Buchanan and Kafoglis' nonreciprocal examples: the assumption that when one individual uses an input an external economy is created for the other, when the reverse is not true, suggests that one individual is a more efficient producer of output (in this case immunity) than the other. When an ideal collective arrangement is obtained, more of the input is given to the more efficient producer, so that more output is obtained per unit of input. The result is that expenditure on inputs (immunization shots) may decrease when collectivization brings about the use of a more economical productive process, but that output (immunity from disease) increases.

Our purpose is *not* to suggest that Buchanan and Kafoglis or Baumol failed to understand this point. Indeed, there are textual passages and footnotes in the two papers we have discussed, and in other works by these authors, which suggest that they did.[16] Our principal purpose is

[16] See the perceptive footnote 9 on page 364 of Baumol, and the paragraph on page 406 of Buchanan-Kafoglis, where they refer to the need to distinguish units of resources supplied from units of final output. See also the previously cited article by James Buchanan and Gordon Tullock on "Public and Private Interaction under Reciprocal Externality." On page 61 of that paper, the authors say:

"It is possible that the reciprocal external economies are such that only through collective organization can a sufficiently *unequal* distribution of services be implemented, so as to achieve efficiency. In this instance, collectivization allows a specific technological improvement in distribution to be made that is not possible under private provision. This seems reasonably common with some real-world public services. Police do not patrol every street with the same intensity. The national defense establishment concentrates its services at strategic points."

rather to clarify the matter and show its essential simplicity. This seems necessary, for many altogether capable economists of our acquaintance who have read the Buchanan-Kafoglis and Baumol contributions have misunderstood them, and have in particular totally failed to see the simple point we have explained above. Another of our purposes is to show that the value of Baumol's contribution is not in any explanation of the Buchanan-Kafoglis examples (it does not by itself provide a sufficient explanation of these examples), but rather in the more general discoveries that it contains. Our final purpose has been to make the true policy implications of the Buchanan-Kafoglis examples clear. Buchanan and Kafoglis' paper often leaves the impression that it calls into question "the theory of economic policy upon which arguments for the collectivization of any activity must be based" [17] as well as the "orthodox policy implication" [18] of that theory. Whatever Buchanan and Kafoglis thought about the policy implications of their argument, they nevertheless do *not* point out anywhere in their paper that a primary policy implication that may be drawn from their examples is that the shortcomings of independent adjustment in a free market with externalities are greater than has previously been understood. There is now a new and additional reason for collectivizing economic activities with externalities. For the Buchanan-Kafoglis finding that the expenditure on a good with externalities may be as great or greater with independent adjustment than under ideal collective arrangements clearly shows that, *when external economies are present, the market mechanism may not only provide less of the good than is optimal (as Pigou told us long ago), but will often also make society pay more for the suboptimal quantity of the good than it would have had to pay for the larger amount an ideal collective arrangement could provide (as we have now discovered)*. Or, to put the same point in a different way, the examples in Buchanan and Kafoglis show that independent adjustment in a market with externalities will normally lead not only to the production of the wrong amount of a good, but also to an unnecessarily inefficient location or method of production. Only in the highly special case where all parties who create externalities provide the same amount of spillover per unit of input will the location or method of production be the same under

[17] Buchanan and Kafoglis, "Public Goods Supply," p. 403.
[18] *Ibid.*, p. 403.

independent market adjustment and under an ideal collective arrangement; and in this special case the traditional Pigovian analysis is appropriate. The traditional Pigovian model overlooks the general connection between externalities and the method of production, and is to that extent inadequate, but the traditional argument that externalities will tend to lead to inefficient market outcomes is now twice blessed.[19]

The considerable practical importance of this matter is illustrated by the police and medical examples cited by Buchanan and Kafoglis. We know that they argued that the total expenditures on firearms, guards, locks, and other forms of protection might well increase if the public provision of police services were to stop. This seems plausible, and we can now see why it is. The private, decentralized provision of police services would presumably be less efficient than centralized provision of the same services, with the result that society might have to pay more for protection even though it obtained less. The same thing might be said of defense, though here the relative inefficiency of private, decentralized defense provision might be so great that even the smallest useful unit of defense protection would be prohibitively expensive for an individual citizen. It was an attempt to "explain" why expenditures on medical care have increased more rapidly in the United States than in Great Britain since nationalization which Buchanan and Kafoglis tell us prompted their study. The explanation of this phenomenon which Buchanan and Kafoglis' various examples suggest is that the production of medical care in Britain is, because of nationalization, more efficient than in the United States, and that the British are getting more health protection than a comparison of expenditures on medical services would suggest.[20] We may personally doubt that this is anything like a satis-

[19] There is of course no suggestion here that all activities with externalities ought to be collectivized, or that there should in practice necessarily be any increase in collectivization. The ideal collectivization assumed here for purposes of analysis and comparison will rarely if ever be attainable in the real world, and in some cases the imperfections of a market mechanism distorted by externalities will be less serious than those of a collective arrangement.

[20] We are thankful to Frederic M. Scherer for calling our attention to the fact that monetary expenditures on the National Health Service may very well understate the increase in the amount of resources allocated to medical purposes in the United Kingdom since that industry was collectivized, for it is often observed that the length of queues has increased since collectivization. The longer queues have an obvious opportunity cost, and may bring a fuller utilization of doctors and hospitals that would reduce the money cost of a given amount of medical treatment.

factory explanation of this very complex phenomenon, but we do emphasize that this explanation is what is implicit in the Buchanan-Kafoglis argument.[21]

V

There is, however, one numerical example in the Buchanan-Kafoglis paper that our analysis does not at first appear to explain. This is a fascinating hypothetical example of reciprocal externalities; that is, an example of a situation in which both A and B confer external economies on each other. The example is in the form of a payoff matrix showing the returns to A and B from different levels of immunization; this example proves that the *expenditure* on a good with externalities cannot only be greater under independent adjustment than under an ideal collective arrangement but also that *output* can be greater under independent adjustment. According to Buchanan and Kafoglis, this finding "suggests that, under independent adjustment, the consumption of final output is overextended." [22] This result seems impossible both in terms of traditional analysis and of the argument of the present paper. Thus we must ask how the logical possibility that has been demonstrated by example can be explained.

The explanation again centers on the differing efficiencies of the different ways in which external economies or collective goods can be generated or produced, but now it will be necessary to distinguish between different degrees of efficiency in the production of a given externality in an over-all or total sense, and different degrees of efficiency in the production of particular marginal units. Assume that in a given case there are two relevant alternative ways to produce a given collective good or externality. One is the manner that results with independent adjustment, which we shall call "means 1." The other way involves con-

[21] After this paper was written, we read James M. Buchanan, *The Inconsistencies of the National Health Service*, London, The Institute of Economic Affairs, Occasional Paper No. 7, 1965. In an appendix to this most challenging pamphlet, Buchanan relates his paper with Kafoglis to the National Health Service in more detail, and explicitly points out that it is logically conceivable that collectivization could have brought increased efficiency which could account for the relatively modest increases in medical expenditures in Britain since the National Health Service was created.

[22] Buchanan and Kafoglis, "Public Goods Supply," p. 411.

centrating all of the generation of production of the collective good in the unit able to produce it at lowest total cost. This latter manner of production, which would be used by an ideal collective arrangement, we shall call "means 2." Though it is assumed that (for any and every output) total costs of production are less by using the second manner of production rather than the first, it is nevertheless possible that, over some limited range, the *marginal* costs of producing the good might be less if means 1 had been used. This is illustrated in Figure 3, where it is apparent that everywhere total costs of production are less using cost curve TC_2, and that output OQ_2 is Pareto-optimal. Nonetheless, when the cost curve embodying less efficient means 1 is used, marginal costs are lower over a considerable (and crucial) range. With this cost curve, total gains are at their maximum at OQ_1 of output—that is at a level of output greater than the Pareto-optimal level of OQ_2. This shows that in certain special circumstances the most advantageous output of a good will be

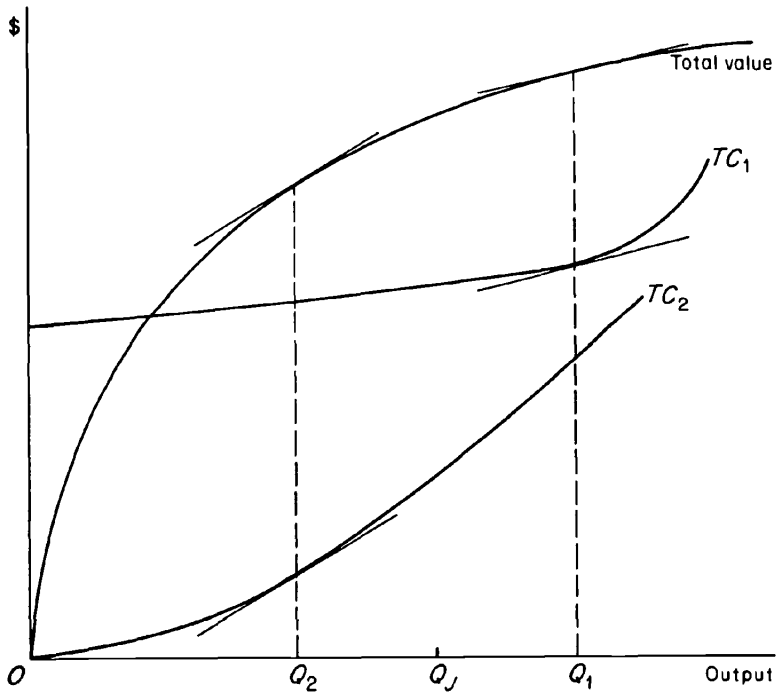
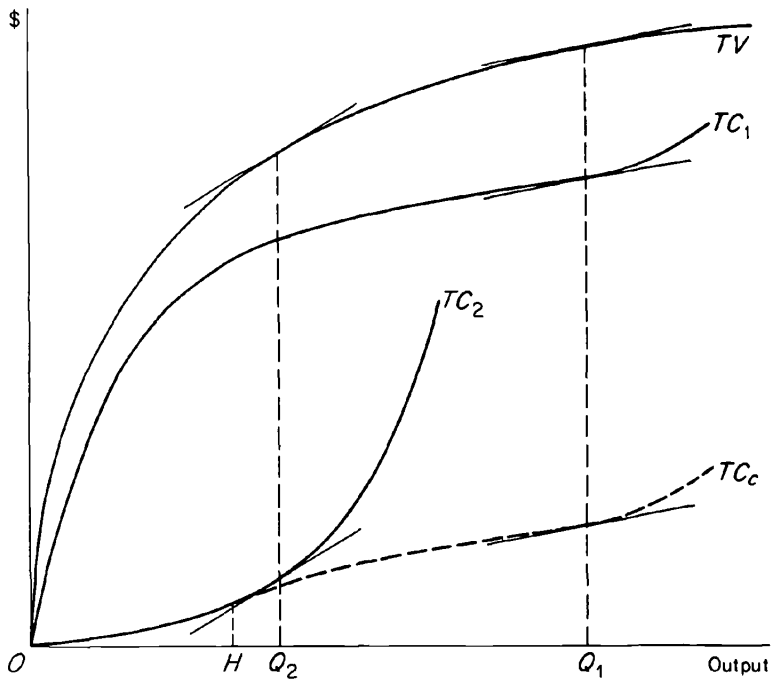


Figure 3

larger when a less efficient means of production is used than it is when a more efficient means of production is used.[23]

[23] The argument in the text required the assumption that an ideal collectivity could not have produced every unit of the collective good at a *marginal* cost as low or lower than that confronting any individual member. In a different case, such as the one described in the figure below, where an ideal collectivity would use means 2 to produce all of the units which could be produced with the lowest



marginal cost by that means, and at the same time enjoy that segment of TC_1 which indicates lower marginal costs over a range with means 1, it would have a cost curve such as TC_c. This cost curve consists of cost curve TC_2 up to point H and has the same slope as cost curve TC_1 after that. If with ideal collectivization it is possible to produce each unit of a good by whatever means has the lowest marginal cost curve for that unit, then obviously no means of production can have a lower marginal cost at any point than can be achieved under ideal collectivization. When this is true, a result of the kind which Buchanan and Kafoglis demonstrated by example and which we explained in the text is impossible. Since, in the case at hand, the profits using the independent adjustment cost curve TC_1 are higher at OQ_2 than at OQ_1, and since total costs of production cannot rise more between OQ_2 and OQ_1 along TC_c than along TC_1 (because TC_c has by definition at every output level a slope equal to the flatter of the two cost curves), it follows that the total gains under the ideal collective ar-

In the presence of external economies, independent adjustment would not, to be sure, lead to output OQ_1. For obvious Pigovian reasons, each party will normally neglect the benefits its production confers upon the others and produce too little of the collective good. In other words, independent adjustment would not attain even the local maximum along cost curve TC_1. Still, it is possible that independent adjustment along cost curve TC_1 could lead to an output such as (say) OQ_j, which is larger than the Pareto-optimal output, OQ_2. We now have an explanation of how it was possible that the Buchanan-Kafoglis numerical example could show a larger *output* of immunization with independent adjustment than with a Pareto-optimal collective arrangement. Their example is in fact accounted for by relationships of the kind we have just described.[24]

rangement must also be greater at OQ_2 than at OQ_1. More generally, no other available cost curve could lead to a higher optimal output than that which would be optimal with the ideal collective cost curve. In such circumstances, it is not possible that independent adjustment could lead to a larger output than is Pareto-optimal.

The assumption which is necessary to the Buchanan-Kafoglis example explained in the text is not, however, always altogether without realism. If the total costs of producing $X + 1$ units of a collective good are less by having individual B produce every one of these units, but the *marginal* cost of the $X + 1$ *th* unit would have been less through individual A *if* he also produced all (or some) of the X units, the assumption implicit in the Buchanan-Kafoglis example would be realistic. In other words, if the marginal costs of generating an external economy through a given individual are a function of the number of units he has previously generated (as is surely often the case), then an ideal collectivity often could *not* produce every unit that it produced with marginal costs as low or lower than would prevail under every alternative means of production. And whenever the *marginal* costs of generating additional units of an externality are higher, over some range, for the ideal collectivity than for some individual member, as they are in the Buchanan-Kafoglis example, a result of the kind explained in the text is possible.

[24] The payoff matrix in the Buchanan-Kafoglis example reveals that the efficiency with which immunity is produced varies between A and B. The Pareto-optimal solution consists in giving three shots to B and none to A, for B is the more efficient producer of immunity. The independent adjustment equilibrium is found when A and B get two shots each, or a total of one more than in the Pareto-optimal solution. Though dividing the immunization shots between A and B means generally higher costs of immunity than would result from giving all of the shots to B, once three shots have been given to B a fourth shot brings so little extra immunity that it is not worth its cost. But if the less efficient practice of having both A and B take immunization shots has been followed, a fourth shot turns out to *add* a considerable amount of immunity. Thus though the total cost of immunity is lower if all of the shots are given to B, over one crucial range the marginal costs of immunity are lower if the shots are divided between A and B in the fashion that would

The finding that output under independent adjustment can be greater than the ideal collective or Pareto-optimal level does not however suggest, as Buchanan and Kafoglis claim, "that, under independent adjustment, the consumption of final output is overextended" even in the special cases we have considered. For if the production processes used under independent adjustment continue to be used, every one can be made better off by increasing output still farther. This is because of the classic Pigovian reason that the individuals are not rewarded for the external economies they provide to others, and will normally produce too little with any *given* method of production. The output is larger than that which would be provided by an ideal collective arrangement because of the special relationship of the different cost functions used in each case. Still, the output under independent adjustment cannot be said to be overextended, since if it were reduced and nothing else were changed, the situation would be worse for all concerned.

VI

Now the apparent contradiction between the alternative collective goods and externalities theories with which we began this paper can be resolved. We started out by pointing to the fact that different collective goods or externality theories were apparently contradictory and that these contradictions had to be resolved before the international implications of the fact that defense was a collective good could be understood. We were concerned about the fact that the present authors' earlier collective goods model and the traditional Pigovian externalities model, both of which indicated that external economies, or more generally collective goods, led to suboptimal outputs under independent adjustment, were apparently contradicted by the discoveries of Buchanan-Kafoglis and Baumol, which appeared to suggest that independent adjustment might not systematically lead to suboptimality.

We are now in a position to see that the conflicts between the different models tend to disappear when they are examined more closely. Though it may have seemed from Buchanan and Kafoglis' discus-

result from independent adjustment. The irregular quality of the payoff matrix is further reflected in the fact that it has two local maxima. One (inferior) local maximum occurs not far from where the independent adjustment equilibrium is found. The other (superior) maximum is of course at the Pareto-optimal point.

sion of their hypothetical and practical examples, which appeared to destroy the orthodox externality analysis, and from Baumol's proofs, which appeared to confirm the Buchanan-Kafoglis findings, that the orthodox analysis was fundamentally incorrect, this was a mistaken impression. In fact, the Buchanan-Kafoglis examples can be explained *entirely* in terms of the fact that the efficiency with which external economies are generated depends on who generates them. When this is recognized, it is clear that the Pigovian case for collectivizing activities with externalities is (contrary to our original impression) strongly reinforced, for independent adjustment in a market will not (except by accident) lead to the use of the most efficient method of generating externalities. Since collectivization can lead to the most efficient organization of the production of the external economy, we have in fact a new case for collectivization, and one that complements the traditional Pigovian argument.

Baumol's striking discovery of the upward-move-that-takes-one-farther-from-the-top does, it is true, show that with incomplete information producers of a good with external economies could in certain disequilibrium situations produce a supraoptimal supply of such a good. But the possibility Baumol discovered could lead to suboptimality as well as to supraoptimality and does not contradict the view that independent adjustment in the presence of external economies systematically leads to suboptimal provision. Moreover, Baumol's finding on the upward-move-that-takes-one-farther-from-the-top does not properly explain any of the Buchanan-Kafoglis examples, and is in fact instead a discovery of sweeping significance that has no unique application to externalities.

Our view in short is that the contributions in the pathbreaking papers by Buchanan-Kafoglis and Baumol are not what they appeared to be, and that they have not impaired, but rather helped to give a new dimension and importance to the orthodox conclusion that independent adjustment in a market cannot deal efficiently with externalities.[25]

[25] Since all of the Buchanan-Kafoglis and most of the Baumol arguments were expressed in terms of external economies rather than diseconomies, we have also confined our attention to external economies. But all of the arguments that have been made can easily be made to apply to diseconomies. Our basic argument would suggest that when the firms producing some good or service generate external diseconomies, independent adjustment will not only mean that too much of this good or service will normally be produced, but also

VII

The principal insight uncovered by our analysis of the recent theoretical advances in this area has a special relevance to defense and military alliances. As we have pointed out before, each ally's military spending provides an external economy to the other allies. And defense is a collective good that can be produced more efficiently by some nations than others. If, say, Germany and Italy are in the same alliance, and it happens that one of these nations is relatively efficient in military matters and the other is not, it would be a necessary condition of economic (and military) efficiency that the militarily more efficient nation provide a larger share of the alliance's military capacity and the other ally export more private goods to it in return. The nation which has the comparative advantage in the collective good of defense would have to specialize in that, and the nation that has the comparative advantage in private goods would have to specialize in such goods. This would make it possible for them to "trade" to their mutual advantage.

Defense in turn can be divided into missile capability, naval capability, infantry capability, and so on. Presumably different allied nations will produce each of these capabilities with different degrees of efficiency. This suggests that there can also be mutually advantageous "trade" between different allied nations in particular military collective goods. In the North Atlantic Treaty Organization, this might mean that the United States would provide the missiles, Britain the navy, Turkey

that the distribution of output among the firms in the industry and the nature of any efforts to minimize the injury inflicted by the diseconomies, will tend to be inefficient. An optimal solution would require that those firms with the "largest" external diseconomies, that is, those that inflict the most injury to others per dollar of output, must restrict their output more than other firms in the industry. More generally, our arguments suggests that Pareto-optimality would require that the least-cost method of dealing with the diseconomy be chosen. And this would often require that part of the effort to deal with the diseconomy be undertaken by those who are injured by it, since the cost involved in their "fencing out" or otherwise avoiding the external diseconomy might be less than the costs involved in preventing or limiting the extent of the diseconomy at its source. This in turn means that the common suggestions to the effect that those who generate external diseconomies ought to be made to pay their victims the full value of any losses they suffer, can work against the achievement of Pareto-optimality. When such a suggestion is put into practice, those injured by the externality have no incentive to attempt to protect themselves against the external diseconomy, even if this would be more economical than dealing with it solely through those who generate the diseconomy.

the infantry, and so on. Each nation would produce that collective good (if any) which it produced with comparative advantage, and its allies would consume the spillover.

It will of course very often be the case that nations (sometimes with good reason) will value military self-sufficiency more highly than the gains in economic and military efficiency which international specialization in collective goods could bring about, just as they often sacrifice the gains from free trade in private goods in order to enjoy tariff protection. But to say that it will sometimes be politically impossible, or on balance undesirable, to follow the dictates of comparative advantage, is not to say that an understanding of comparative advantage is useless, either with respect to collective goods or individual goods. Thus, however great the difficulties and complexities of the real world may be, there is still surely a need to call attention, in a relatively general, theoretical way, to the potential gains from trade in collective goods.[26]

In some particular, practical cases these gains have been subjected to penetrating study. Malcolm Hoag pointed out some time ago, in an important article, that the NATO alliance could be made more efficient by applying the principle of specialization according to comparative advantage to various types of military capability.[27] Hoag did not use the theory of collective goods in his argument, but relied instead on the theory of international trade.

The point that Hoag and others have illustrated in certain concrete

[26] This paragraph was inserted after the paper was presented at the conference because the comments helpfully revealed that these points, which we had barely mentioned and to some extent even taken for granted, in our presentation to the conference, should have been made clearly and explicitly.

We have also not distinguished (in this particular paper) the collective (alliance) and individual (national) benefits which usually accrue from military expenditures by an alliance member. This does not mean, however, that our analysis does not hold when an alliance member receives significant private or purely national benefits from its defense spending. This individual benefit would be reflected in a higher demand curve for its own as opposed to its allies' defense spending. When there are significant noncollective or purely national benefits of this sort within an alliance, even inefficient producers of defense (or of some particular type of defense) may produce some of the good, and this could be rational, not only from the point of view of the individual nation, but sometimes also from the point of view of the alliance as a whole. The rationale for such expenditure is that the private benefits which result from it more than compensate for the fact that it is more expensive than the military capability that could be provided by the more efficient allies.

[27] Malcolm Hoag, "Economic Problems of Alliance," *Journal of Political Economy*, December 1957, pp. 522–534.

cases has, however, been ignored in all of the theoretical literature with which we are familiar. This is the point that collective goods or externalities imply spillovers from one party to another, and that these spillovers should be regarded as a special type of trade.[28] As we inferred from the Buchanan-Kafoglis examples, these spillovers will normally be produced with different degrees of efficiency by alternative producers. This means that economic efficiency will usually require that some produce one kind of spillover, others another, and that still others produce only private goods. In brief, the theory of externalities and collective goods has to be combined with the theory of comparative advantage. That is the most basic and general point of this paper.

When the theory of collective goods and the theory of comparative advantage are combined, the problems of military alliances can be studied in a more unified and coherent way than has been possible before. For when these two theories are used together, it becomes evident that an alliance organized in the way that NATO is organized suffers from at least two systematic inefficiencies. On the one hand, it will normally

[28] When each of the members of a group produces some *single* collective good that also benefits the other members of the group and they thus all trade spillovers of this good, there is the question of how they should compensate each other to obtain a Pareto-optimal supply of the good. If each member of the group is allowed to choose his own output of the good, there will need to be transfers of money or other sidepayments, since otherwise each member will ignore the benefits of his production to others when deciding how much to produce. This is not in dispute.

But it is perhaps not generally known that some obvious and seemingly ideal compensation schemes that economists might suggest will in fact *not* be consistent with Pareto-optimality. Take, for example, this compensation rule: "Each member pays to the other members the value to himself of the spillover they provide." This intuitively appealing compensation scheme would in fact lead to a *supra*optimal supply of the collective good. This is not the place for a complete, lengthy explanation, so that will be left for a future publication. But in essence, the reasoning is as follows. When a group member decides whether or not to produce an additional amount of the collective good under this compensation scheme, his return can be distinguished into three parts: 1) the value of the additional amount of the collective good to himself, 2) the compensation he receives from the other members (which is under the rule equal to the value of the additional amount of the good to the other members), and 3) the *decrease* in the amount of compensation he must pay the other members for the collective good he has received from them (this will tend to decline as his own output increases, assuming declining marginal utility of the good, for the more he produces, the less valuable to him is any given amount of spillover from others, and the less the compensation he owes). Since returns 1) and 2) are together equal to the full social value from the additional amount of the collective good, the additional return of type 3) leads to a supraoptimal production of the collective good for the group.

produce a suboptimal output of alliance defense, and on the other, it will fail to produce that amount of defense it does obtain in the most efficient places. The potential gains from greater integration or collectivization therefore seem considerable. But among nations, as within them, there is also a reasonable concern for the sovereignty of the individual decision-making unit.

COMMENTS

PAUL G. CLARK, Williams College

The central issue raised by Hoag's paper is: Where does analytical marginalism *not* apply in the economics of defense? The particular basis for exceptions which he argues is the phenomenon of increasing returns to scale. The main question on which I shall try to comment, therefore, is the significance of exceptions based on increasing returns. At the end I shall also mention briefly a different basis for departing from analytical marginalism.

It is helpful to be clear at the beginning that we are concerned with peacetime decisions about the preferred mix of inputs for defense preparations. Moreover, we are concerned with decisions by the United States—that is, for our present purpose, by a country with many diverse worldwide interests; seeking conservative and reform objectives which call for ability to respond militarily rather than to initiate military hostilities; with a large, high-income, progressive economy; dealing with the Soviet Union in what is still essentially a bipolar nuclear situation; but operating in other respects in a complex system of allies, neutrals, and hostile countries. Our analysis might well be different for other defense decisions or for countries in other circumstances.

Hoag's principal case is derived from combat. It is that there are dramatically increasing returns to scale for defense inputs into a particular military operation, until a margin of superiority over the enemy's inputs has been attained sufficient to win militarily. Accepting this combat phenomenon as valid, let me discuss its implications for defense preparations at three levels—the engagement, the theater, and the world.

At the level of the engagement, I presume it is generally accepted that, with a given budget for the engagement which provides this margin of superiority, there are a wide variety of input mixes with decreasing returns for particular defense inputs. The question is whether the phe-

nomenon of increasing returns to the budget for the engagement, up to this point, calls for a significant departure from the marginal approach. This question, however, must be considered at the level of the theater.

Within the theater, there are many possible engagements which potential enemies may choose, and accordingly the model for deciding on defense preparations must be one with multiple contingencies. Accepting the phenomenon of increasing returns to defense inputs specialized for particular engagements, let us assume further that the given budget for the theater is not sufficient to provide in this way the desired margin of superiority for all contingencies. The preferred adjustment here, however, is not to abandon some contingencies for others, which would leave potential enemies free to choose engagements so as to make the returns to the theater budget zero. The preferred adjustment is rather to provide some mobile defense inputs which, together with some specialized inputs, will give similar expected outcomes in all contingencies. Then quite possibly, depending on the substitutability among inputs permitted by military technology, and on the inventiveness of the decision-makers, the situation may be converted into one in which the given theater budget is operating in a range of decreasing returns, now positive, while particular inputs within the preferred mix are also subject to decreasing returns. It is an empirical question, but I suggest that this is a common rather than a rare situation in theater defense preparations. If after such adjustments there are still increasing returns to the budget for the theater, we must move to the level of the world.

Within the world, there are many possible theaters which potential enemies may choose, and we can repeat the same kind of analysis. Again, by adjusting to a suitable mix of mobile and specialized inputs, a given budget for the world may quite possibly be converted from the range of increasing to a range of decreasing returns. There are two further possibilities as well. The budget for worldwide preparations can be expanded to get over the range of increasing returns; this has been an especially suitable option for the United States because of her large, high-income, progressive economy and similar allies. Moreover, since theaters within the world are much less interdependent than engagements within a theater, defense of some theaters can be abandoned to attain the range of increasing returns in others; actually it does not appear that this option has been necessary for the United States and her allies.

Hoag also presents two other cases of increasing returns which he suggests call for departures from analytical marginalism. One is economies of scale in production of particular defense inputs. In view of the great variety of defense inputs required, I would suggest that the effect of this phenomenon is simply to bend design and procurement further toward multipurpose, i.e. mobile, rather than specialized inputs. The other is increasing returns over some range for network-type inputs, such as a reasonably reliable antiballistic missile system. For defense of cities, I would suggest that this range is significantly limited by the rapidly decreasing size of cities, when ranked by size, characteristic of industrial countries.

All in all, Hoag's paper should contribute to making defense analysts sensitive to phenomena of increasing returns where they arise in defense problems, and prepared to adjust their analyses accordingly. At the same time, I feel that the cases of increasing returns which Hoag cites do not constitute very extensive grounds for departing from the general approach of analytical marginalism. It is not really clear to me how significant he considers these exceptions based on increasing returns. To recall his phrase, there may be "no inconsistency" here, even though there appears to be "a contrast."

On the other hand, there is a crucial field of defense decisions in which a quite different ground for departing from the marginal approach exists. Hoag clearly recognizes this other ground and simply has not taken it as the subject of his present paper. I refer to decisions on the major types of military capability to obtain, such as NATO conventional forces, Polaris submarines, or an antiballistic missile system, in which a central element of the problem is the reaction of the Soviet Union in her decisions on major types of military capability. In this bipolar field, a game analysis which is intrinsically nonmarginal is called for. It seems to me that this is the most significant answer to the general question raised by Hoag, about where analytical marginalism does not apply in the economics of defense.

Let me add only one comment on the paper by Olson and Zeckhauser. It presents a persuasive rebuttal to criticisms of the generally accepted proposition that collective goods, in this case defense services of allies, tend to be provided at less than Pareto-optimal levels unless appropriate nonmarket institutions are introduced. Here the scale of a country's contribution is another decision for which the marginal approach is in-

appropriate. It is interesting to note that the basis for the exception is again the game character of approaches toward the optimum—though in this case it is a game among allies.

RICHARD N. COOPER, U.S. State Department and Yale University

I will ignore Lanchester's precept to concentrate one's forces in local engagement by dividing my comments between the two preceding papers. I take comfort, however, from the fact that I am well flanked by my fellow discussants.

The basic point of Hoag's paper seems to be that the skepticism by military officers regarding the application of marginal cost-benefit analysis to military problems is justified when one takes into account Lanchester's Law and other manifestations of increasing returns to scale. The attempt to solve problems of national defense within the conceptual framework of economics, pursued most notably and successfully by RAND economists in the midfifties, is given the caution signal. While I agree with many of Hoag's observations, I would temper what I take to be his principal observation in three ways.

First, increasing returns to scale in questions of national defense may not be present so frequently as Hoag implies. It is useful to consider four stages of the input-output sequence related to national defense: (1) the cost to the national economy of raising and equipping its forces, developing new weapons and weapons systems, and producing military hardware; (2) the deployment of military forces so raised and equipped; (3) the success such forces obtain in achieving military objectives, e.g., beating an enemy force or preventing an enemy attack; and (4) the social utility of these military successes to the nation. Typical discussions of military tactics and strategy focus only on the second and third of these steps. Lanchester's Law is couched in these terms, and most military men operate within this frame of reference. The first and last steps have been left largely to "politics." While they are inevitably political in the broadest sense, the economist can usefully apply his trade to the whole process.

Increasing returns of the type considered by Hoag apply largely to the probability that military success will attend various deployments of

NOTE: The views expressed here do not necessarily represent those of the U.S. State Department.

military forces. But increasing returns at this stage can be overshadowed by the rising economic and social cost of increasing total military forces (or, at a less global level, the rising opportunity costs in terms of over-all military success of enlarging a *particular* military deployment); or by the declining marginal utility of the additional benefits as one gets into the area of "overkill," to apply that expression to local and conventional engagements. Both factors will limit the desirability of expanding indefinitely the particular deployment. This sort of consideration presumably led to Hoag's choice of the 8 to 5 ratio of British to German ships in the pre-1914 Atlantic, rather than 10 to 5 or 12 to 5 or some higher ratio, even though the increase in probability of a local German defeat might continue to rise as the ratio is increased.

Second, even when increasing returns do prevail throughout the total four-step range for decision-making, marginal analysis is far from irrelevant. When increasing returns are present, marginal analysis should indicate that an additional unit of input will bring an enlarged marginal benefit. In that case, use of the input should be expanded. The expansion should stop only when marginal costs equal marginal benefits, regardless of whether both are falling or both are rising. Needless to say, this equality will be reached only after marginal benefits have begun to fall faster than marginal costs, or marginal costs have begun to rise faster than marginal benefits. But in military as in economic affairs, all good things must come to an end, and sooner or later increasing returns in one part of the process bearing on decisions will be swamped by rising costs or decreasing returns in others. Consumer saturation has its military analogue in overkill.

Third, while marginal analysis admittedly does fail in the presence of increasing returns to indicate whether an enterprise should be undertaken at all (not, if it is undertaken, how far to expand it), in defense economics the question whether to undertake an enterprise or exercise at all does not depend alone, or even primarily, on static analysis of increasing returns. It can be answered only by taking into account the adversary-relationship—the dynamics of an arms race or a force-mobility race or an information-gathering race. This relationship is not present in the decreasing-cost generation of electricity for atomistic consumers. But it dominates the analytics of military confrontation. The enemy will typically understand the same principles of increasing returns to concentration of forces, mobility, and static defense. Germany often gained

great advantage during World War II by concentrating its forces for attack and by avoiding heavily defended areas. Accurate information takes on critical importance in achieving this concentration, as in the Battle of Midway in which the United States gained local sea and air superiority, despite its over-all inferiority, by cracking the Japanese code; or as at Pearl Harbor where the Japanese caught a superior force off guard. Gaining local superiority—in alertness and effectiveness as well as in numbers—requires outwitting the opponent. But here it is bilateral duopoly and the presence of uncertainty or ignorance which complicates the analysis, not the existence of increasing returns implied in a Lanchester's Law.

Thus, loosely speaking, in "tactics" marginal analysis is helpful even in the presence of increasing returns, while in "strategy" the existence of increasing returns is overshadowed by the dynamics of interaction.

The foregoing remarks, like Hoag's paper, concentrate on Lanchester's Law and its variants as the principal example of economies of scale in military situations. Hoag refers only briefly to economies of scale in production. Perhaps that is the correct relative weight to attach to these two areas, but I would have preferred to see more discussion of the latter. Standardization of weapons permits both production and logistical economies. With the development of satellites, gathering weather and other photographic information involves considerable economies of scale; and crytography is perhaps the most interesting case of all, where a myriad of small and seemingly insignificant pieces of information is often needed to make over-all sense out of the enemy's behavior and a very large establishment is required for maximum advantage. The economies in these critical areas are enormous and the benefits are enhanced precisely because of the possible substitution of information, surprise, and maneuver for sheer force, which Hoag rightly emphasizes.

The presence of economies of scale or, more correctly, of discontinuities or lumpiness in military operations suggests that countries large in economic resources will gain increasing advantage over small countries or over countries unwilling to devote their resources to military and defense-related expenditures. Even medium-sized countries apparently cannot develop today's weapons systems alone, as evidenced by the British decision to buy the F-111 from the United States rather than develop its own version. The British-French effort to maintain a modern aircraft industry through joint development of the Concorde, the

French reliance on U.S. firms for a computer industry, Euratom, and increasing discussion of a European satellite system (including launching vehicles), all indicate that alliances are sought to overcome these discontinuities.

The Olson-Zeckhauser paper is concerned with alliances, but its focus is on public goods rather than economies of scale. Most of their paper is occupied by criticism of the finding by Buchanan and Kafoglis that pure public goods may be purchased in excessive amounts if left entirely to the market. This finding, as Olson and Zeckhauser point out, rests on the selection of an inefficient production technique in the case of market decisions, whereas under collective decision the more efficient production technique (involving externalities) can be chosen. Thus where externalities are present, we are not only likely to get too little of the product in a free market system, but what we get will be produced inefficiently as well—possibly leading to larger expenditures than otherwise but with less product.

This is an important proposition. It is not of course a new justification for intervention in market decisions. Greater efficiency has long been one reason advanced for collectivization of goods and services, whether that efficiency is based on economies of scale, on externalities, or on avoidance of avoidable overhead costs. Both economies of scale and externalities have been used, for example, as an argument for public development of the Snake River—that one large public dam was less costly for the same product than three small private ones. Efficiency based on the presence of quasi-joint products (where producing good A greatly lowers the cost of producing good B)—one version of externalities—has frequently provided the rationale for multipurpose river development, where recreation and navigation are combined with public power. Britain's move to nationalize the health services, which before nationalization were said to have entailed many wasteful overhead expenditures and to have provided too little service per unit of cost, was based in part on the social disutility of having substantial numbers of the public who either could not or would not afford proper medical care; in part on the desire to avoid avoidable costs. Compulsory social security also combines reduction of a social disutility (a large population of indigent aged) with real economies from pooling all risks, thus avoiding the need for developing many different risk categories and eliminating the opportunity for self-selection. In all three cases reliance on the

market place would not only have failed to take into account the external effects, but also would have resulted in a higher cost per unit of output.

Having agreed with the major proposition of Olson and Zeckhauser, I will quibble with them on two minor points and on their application of the finding to alliances.

First, the authors implicitly treat the term "collective good" as coextensive with "externality." I would consider externalities to represent a much broader class of phenomena than collective goods. Externalities are essentially nonappropriable by-products of some producing or consuming firm or person. The fragrance of an orchard and the smoke of a factory are examples. In contrast, a collective good is itself the object of a decision to consume, and all consumers enjoy the same product.

A collective good can be regarded as a limiting case of an externality. Thus when a wealthy neighbor hires a private guard to protect his property, some protection is presumably also afforded to the immediate neighborhood insofar as potential housebreakers and other criminals are not likely to perform their activities within view of the guard even when his responsibility, strictly speaking, does not reach out to the property in question. There is an external benefit. If the residents of a city block decide to hire a local police force, each member of the block will be fully protected, even though he does not participate in financing the force (provided the crime deterrent effect of the police force is complete). Crime deterrence is a pure collective good. Thus there is a continuum. As the externalities become more and more important relative to the principal item of production or consumption, more and more of a "collective good" aspect is involved. Pure public goods represent the limiting case in which the externalities dominate entirely and the "principal" good being produced or consumed disappears—or rather becomes coextensive with the "externality." [1]

Second, I am troubled by Olson and Zeckhauser's apparent endorse-

[1] Viewing collective goods merely as a limiting case of externalities ignores one important distinction which is sometimes made. An externality can be defined to involve the nonappropriability of costs or benefits through the market mechanism. Public goods or collective goods, on the other hand, can be defined to include all those products for which consumption by A in no way reduces consumption by B, C, etc., even though exclusion from the benefits (appropriability) is technically feasible. An electronic "lighthouse" with coded signals and pay TV are usually given as examples of collective goods of this type. But in each of these cases creating exclusion incurs additional costs.

ment of the point made in Baumol's paper, namely, that "the usual marginal rules for maximization can lead the society *away from* the peak of the social welfare function." It is not clear what the "usual rules" for maximization are when more than one instrument variable is involved. It is true that guidance by a single partial derivative can increase the horizontal distance from the peak of a social welfare function even though it increases welfare, as Olson and Zeckhauser illustrate. But if the social welfare function is convex—i.e. it has no wrinkles, valleys, saddle points, etc.—application of the gradient method will always lead one both up the hill and toward the ultimate objective, never away from it. The gradient method selects the weights to be attached to the various action variables, the weights being chosen so as to maximize the rate of climb on each step up the social welfare hill. Using it, we do the best we can on each little step, when we are not sure where the peak is. (If we do know exactly where the peak is and can take big enough steps, we should not consider "marginal" rules in the first place.) This approach assumes implicitly that costs of horizontal movement are uniform. If this is not so, a different rule of maximization would be appropriate, but this would also call for a different definition of "movement toward the peak."

Finally, while I agree with the basic observation of Olson and Zeckhauser that collective goods involve an important dimension of efficiency, and that efficiency in production may vary substantially among those who could provide the collective good, I find their ready application of this observation to military alliances too uncritical and unqualified. There are several important reasons why members of an alliance may not want to divide their responsibilities in the provision of collective goods strictly along the lines of "comparative advantage."

A defensive alliance, like a private police force, certainly has a public good aspect, as the authors have shown in another paper; but it is not a pure public good. If strategic deterrence were effective and could be relied upon to be 100 per cent effective, it would be a pure public good. But if deterrence cannot be fully counted upon, then the strategic alliance ceases to offer a *pure* public good. Military defense runs into capacity limitations, and not all areas can be defended equally well. Choices must be made in deploying forces in battle, and those citizens whose territory is left undefended will lose their "share" of the alleged public good. Consumption by A does reduce consumption by B.

This alone—the possibility that deterrence will fail—offers one power-

ful reason why alliance members would not want to divide their responsibilities and contributions strictly along the lines of "comparative advantage." If deterrence fails, their interests begin to diverge.

A second important reason why strict comparative advantage will not offer the optimal pattern of production among members of an alliance is the presence of joint products with some—but not all—military expenditures for the alliance. The presence of military forces in an alliance can serve national foreign policy purposes quite apart from the objectives of the alliance. As long as national authorities maintain some control over their forces or as long as some of their forces can be withdrawn from commitment to alliance objectives, they will have national value beyond their contribution to the alliance. But the possibilities for nonalliance use of military forces vary greatly by type of force.

Provision for internal security is an obvious nonalliance use deriving from some weapon systems. The rationale for a land army or for armed coastal ships is defense of the territory; but these same forces double as police reserves. Such uses of military forces is frequently made around the world, and is occasionally used even in such stable societies as the United States and Britain. Polaris missiles are not helpful in curbing riots in Little Rock or Watts, but infantry reserves are.

A joint product deriving from other types of military expenditure, admittedly more controversial, concerns the civilian by-products of military research, development, and production of continually improving weapons systems. Keeping up-to-date militarily through the nation's own efforts serves both to maintain in being a continuously exercised body of scientific and technical research skills, and to provide technological "fallout" to the civilian economy from military research in metallurgy, plastics, aerodynamics, etc. For both reasons the country allegedly remains in the forefront of advancing knowledge technology, and this can contribute both to national prestige and to material well-being.[2]

For all these reasons it is doubtful whether following strict lines of comparative advantage in military expenditures within an alliance is either practicable or optimal. Few countries will want to specialize com-

[2] The controversy, of course, is whether on balance military R and D has a positive external effect as implied above, or a negative one. Some students of the subject have suggested that the demands placed on scientific and technical skills by military requirements, by withdrawing such skills from the civilian economy, more than offset the various benefits to the civilian economy which result from Government expenditures in these areas.

pletely in one military activity, e.g., provision of foot soldiers, no matter how great their comparative advantage in doing so. For much the same kinds of reasons that less-developed countries want to foster and protect manufacturing industry, the members of an alliance will want to foster and protect certain kinds of military expenditure. And the result will not necessarily be suboptimal.[3]

[3] For these same reasons, incidentally, provision of a public good may not fall below the optimum amount when left to private decisions. If *private* joint products are important enough, more than enough public goods will be provided. If candy also immunized against smallpox, we would not need school vaccination programs.

JACQUES VAN YPERSELE DE STRIHOU, International Monetary Fund

In their interesting paper Mancur Olson and Richard Zeckhauser have addressed themselves to the question of whether the market provision of a public good would lead to suboptimal amounts of the goods being produced. Refuting the objections from Buchanan, Kafoglis and Baumol, the authors conclude that when external economies are present, independent adjustment in a market will not only provide less of the good than is optimal, as Pigou had shown, but will often also make society pay more for the suboptimal quantity of the good than it would have had to pay for the larger amount an ideal collective arrangement would provide. Applying this reasoning to the problem of military alliance which produces a public good: defense, Olson and Zeckhauser conclude that an alliance organized in the way NATO is organized suffers from at least two systematic inefficiencies. On the one hand it will produce a suboptimal output of the public good, defense; and on the other, it will fail to produce that amount of defense it does obtain in the most efficient places.

In these few comments [1] I would like to question the application to the military alliance of the principles developed in the first part of the paper. I will first emphasize the limits of the basic assumption underlying the

NOTE: Any views expressed in these comments are those of the author and should not be interpreted as reflecting those of the organization to which the author belongs.

[1] In these comments I will draw on some elements of my Ph.D. Dissertation on "Sharing the Defense Burden among Western Allies," Yale University, 1966, which was written under the helpful guidance of Professors Richard Cooper and Gustav Ranis, and with the financial support of the Council on International Relations at Yale University.

last part of the paper, namely that defense is an international public good in the alliance. Showing that defense on the international level is only partly a public good and pointing out another characteristic of defense, namely that it provides joint products private to each country, has implications for the efficiency with which defense is produced. These implications, I believe, will invalidate partly at least the normative conclusions drawn by the authors of the paper with respect to the degree of specialization which would be optimal for the countries of the alliance.

I will also discuss briefly the fact that the joint-product characteristic of defense may also lead to a lesser degree of suboptimality in the provision of external security.

Let us first analyze briefly the concept of external security, and see to what extent it is a public good on the national level. Then we will analyze how in an alliance it can be considered as an international public good.

Assuming a given threat common to all Western countries, the security of the citizens of any NATO country A involves at least three aspects at a given moment.

1. The deterrence controlled by country A and that provided by its allies to deter the enemy from attacking country A or any other NATO country whose security is directly related to A's security. As is well known, deterrence is based on a force capable of inflicting on an aggressor, even after a massive surprise attack, enough losses to deter him from launching such an attack and on the willingness to use these forces to oppose any aggression. Together the ability and the willingness to use force will make the threat of retaliation credible. Deterrence can be expressed in terms of probability of nonattack by the enemy.

2. If deterrence fails, security then depends on the country's ability with its allies to defend itself and escape destruction. This is what the U.S. Secretary of Defense calls the "damage limitation capability." It can be expressed in number of men saved in cases of enemy attack.

3. A third aspect of security, is the ability each country has, to prevent being drawn into another country's war against its own will.

Let us now recall the two characteristics of a public good: (a) the "nonrivalness" characteristic, or as Samuelson has expressed it, that each individual's consumption of such a good leads to no subtraction from any other individual's consumption; (b) the nonexclusion characteristic

which means that it is not possible or not economical to exclude people who do not pay for the services from the benefits that result.

Let us now analyze how the two first aspects of security can be considered as a public good for the citizens of a given country A. We will then transpose the problem to the international scene.

1. Deterrence on the national level is the example closest to that of a pure public good. The first characteristic of a public good is fully met indeed. The enjoyment of deterrence for instance by a New Yorker does not decrease the security of a Californian. The cost of applying this deterrence to one more individual coming into this country is zero. The second characteristic of nonexclusion is also fully met. No one can be excluded from the benefit of national deterrence once it is provided.

2. The second aspect of security—defense if deterrence fails—does not meet in all cases fully the characteristics of a pure public good and in some cases is even a private good.

Once deterrence fails, one meets a capacity problem, even in the use of the equipment initially designed for deterrence. The means of defense will not necessarily provide security to all the citizens of the country. While many instruments of defense will provide the same degree of protection to the whole country, others will extend their protection only to limited areas and will be public goods only for the citizens of this area, and not for those outside it. Indeed in this case if citizens of one area receive more protection this means fewer resources will be left for the protection of citizens in other areas.

Examples of defense instruments with strictly local impact are community shelters or local antimissile defenses. In some cases even the protection provided will be a private good because consumption of security by one individual will rival that of another. Examples of this are private nuclear shelters which provide a certain external security but which do not meet the first or the second characteristic of the public good, once deterrence fails.

Let us now transpose the problem onto the international level and see to what extent security can be considered as an international public good in the NATO Alliance.

1. Deterrence meets partly at least the first characteristic of a public good. Indeed, assuming for a moment that the alliance is composed of two members, the U.S. and Europe, one may consider that a certain

amount of deterrence against enemy attack on Europe is a part of U.S. security. Since this deterrence for Europe is already part of the security the U.S. provides for itself, the cost of extending this deterrence to Europe is zero. Similarly, the deterrence against attack on the U.S. is part of European security.

However, the deterrence against attack on Europe which is part of the security of the U.S. is not necessarily the same as the amount of deterrence provided to U.S. citizens. Pledging the same deterrence to Europe in all circumstances as to the U.S. may indeed include positive marginal costs and may even decrease the level of U.S. security, if we recall the third aspect of security—the ability to stay out of other countries' conflicts.

The second characteristic of a public good, nonexclusion, is not met. Even in an alliance, the U.S. for instance, could decide to decrease its deterrence for a European power by decreasing the credibility of its defense of this ally. It is partly the awareness of this possibility which has led France to start building its own deterrence.

France believes that the deterrence provided by the U.S. to its allies is not perfectly automatic and not of the same value as the one provided to American citizens, due to different costs and risks. France emphasizes that while the deterrence of attack on the U.S. rests unambiguously upon nuclear retaliation, this is not the case for Europe.

2. The second element of security, defense if deterrence fails, is not a pure public good. The first characteristic of a public good is not necessarily met. If deterrence fails, a capacity problem is met in the ability to defend citizens. So, in some cases providing more defense for European countries may mean less protection for the U.S., although the defense of different countries will not necessarily be rival in all cases. For instance, the use of U.S. missiles to destroy the attack centers of the enemy will also protect any allies who face the same threat. The U.S. Secretary of Defense recently emphasized this aspect by saying ". . . we do not view damage limitation as a question of concern only to the U.S. Our offensive forces cover strategic enemy capabilities to inflict damage on our allies in Europe just as they cover enemy threats to the continent." [2]

The second characteristic of a public good—nonexclusion—is not

[2] Statement to the House Armed Services Committee cited by *The New York Times,* February 19, 1965.

perfectly met either. If deterrence fails, the U.S. for instance is not bound to give the same protection to Europe as to itself.

Since the external security in the alliance is not a pure public good, the member countries will avoid the extreme specialization, which the authors of the paper seem to advocate, and which would not allow them to operate independently—for the case where they would benefit from only a part of their allies' deterrence—or for the case where deterrence fails. In other words, accepting a complete specialization according to comparative advantage will not be optimal for the member countries.

Another characteristic of defense, namely that in addition to alliance security it provides joint products which are not, or are only partly, shared with other allies, also affects the efficiency with which defense is provided in the alliance. Such benefits are, for instance, the internal security that can be provided by the armed forces, or nonalliance political benefits which can be obtained through the possession of a strong military force or the actual use of it.[3] Complete specialization in accordance with comparative advantage may not allow the member countries to obtain the joint products of defense which they also seek. For instance, a country which would have a comparative advantage in providing missiles may not want to limit itself to this category of military equipment but may want to provide part of its contribution in conventional forces which can help it to obtain national joint products such as internal security.

This characteristic of defense—namely that it provides joint products —can also affect the degree of suboptimality with which external security is provided in the alliance. Indeed the existence of these joint products may induce countries to provide more defense than they would provide if there were only the external security benefits. For instance at the time of colonial wars, some Western countries had much larger forces—which could also be used for the common security—than they would have had if defense provided only the common external security.

Let me briefly summarize my comment. While agreeing with the theoretical concepts developed by Olson and Zeckhauser in the first part of their paper, I believe that their application of this theory to the military alliance is not entirely correct. It would be optimal for the

[3] For a more detailed discussion of these joint products, see J. van Ypersele de Strihou, "Sharing the Defense Burden among Western Allies," *Review of Economics and Statistics* (forthcoming).

countries of the alliance to make their contributions according to comparative advantage, if defense were a public good and did not provide private joint products. But, as I have tried to argue, defense in the alliance is not a pure public good, and it provides joint products to each country which are not or are only partly shared with the allies. Finally the joint product characteristic may tend to decrease the degree of suboptimality in the provision of defense in the alliance.