ASSOCIATION FRANCAISE

de SCIENCE POLITIQUE

<u>Théorie politique:</u> Quelques développements récents

OPTIMAL BOUNDARIES AND DECISION-RULES: A LIBERAL THEORY

OF EQUITY AND EFFICIENCY, AND ITS LIMITATIONS

par

Brian BARRY

SECTION 1

Like Professor Boudon, I wish to take as my point of departure a book by an eminent American author. This is <u>Size and Democracy</u> by Professor Robert Dahl and Professor Edward Tufte. 1

Stanford, Calif: Hanford University Press; and London: Oxford University Press, 1974.

Without intending any discourcesy to Tufte I shall refer to Dahl as the author in what follows.

Dahl has two premises in this book. The first is the desirability of democracy, which we may identify roughly with the control of collectively binding decisions in a political unit by the members of that political unit, either directly (by voting on issues and the outcome corresponding to the majority preference), or indirectly (by voting for representatives who decide issues by majority voting.) The second is that it is desirable for the boundaries of units of political decision-making to coincide with 'problems'. A 'problem' occurs when the actions of one person or group affect another person or group adversely: the paradigm offered is the pollution of a lake by those living round it. The two premises are united in the proposition that people should have an opportunity to influence actions by which they are affected. For Dahl suggests that the only effective way in which this object can be achieved is (a) for all those affected by the action to form a unit of political decision-making and (b) for the members of this unit to have an opportunity to influence the decisions that are taken by that unit.

There are many things in Size and Democracy besides these

propositions but I believe that these are central to the book. Dahl himself ends the book (in his 'Epilogue') by emphasizing the

For a discussion of the book as a whole see my review in Government and Opposition, Autumn 1974.

importance of these points. He says that 'democratic ideas as we have known them up to now provide no adequate guide for discovering an answer to the question: how can democracy be maximised in a world unprecedented both in numbers and in the extent of human interdependence? (pages 138-9). The question is how 'the capacity of ordinary human beings to exercise rational control over their lives is to be enhanced' and 'interdependence among vast numbers of human beings is not to foster a miasma of legitimized domination', (page 139). The answer is that 'a more adequate theory (of democracy) would surely move from wings to center stage the neglected problem of political units and their interrelations. If our conclusion is correct that no single unit should be judged as optimal for democracy today... then what units do we need, and how should they be related?' (pages 139-40).

My objective in this paper is to take up the challenge posed by Dahl, in the hope of making some contribution to our understanding of these very difficult and important questions. I shall begin by re-analyzing the political implications of Dahl's fundamental proposition. I shall then argue that there are many highly significant features of political life that cannot be accommodated by the theory developed in this way, and that a quite different kind of theory is required in addition.

At the centre of Dahl's theory is the individual: it is through his eyes that we compare the advertages and disadvantages of organizing social affairs in alternative ways. What he wants is the maximum influence over 'matters that are highly important to him', which means, as I have observed, that he wants 'maximum effectiveness' over the decisions of any political unit in which he is included but he also wants the political unit to be able to 'deal with these matters' that are 'highly important to him', (page 136). The image summaned up here is of a world made up of individuals all scoking, through political association, to gain their ends. Those with a common problem - who are affected by the same set of actual or resaible actions - form a political unit to deal with it. Thus, those who live round a lake form a political unit to deal with pollution, and so on ad infinitum.

Now Dahl, it should be said at once, rejects such a kaleidoscope of political authorities, but it is important to see that he does so on strictly practical grounds. As the first and the last sentences I shall quote from him shows, he accepts that the logic of his position is that ideally political units should be created ad hee so as to coincide with the 'problem' with which they are to deal. 'If boundaries can be too small or too large, depending on the problem in hand, it might be thought that a theoretical solution would perhaps he found with a system having an indefinite number of units without permanently fixed boundaries, a system capable therefore of ready and infinite adaptability. But 'the costs of communication and information, and therefore of control, would become overwhelming if citizens were confronted with an indefinite number of changing units.' Therefore, 'one task of democratic theory may be to specify not an optimal unit but an optimal number of units with comparatively fixed boundaries.

The boundaries of each unit would be too small or too large for all the functions assigned to it; but the costs of a small number of units with relatively fixed boundaries would be less than the costs of any larger number of units, or of constantly shifting boundaries,' (page 141). The crucial point to hang on to is that the criterion for 'too large' and 'too small' is still given by the number of people 'concerned' in an issue, and the concession made is to practicability.

SECTION II

It will be illuminating, I think, to begin by discarding practical difficulties about costs of communication, information and control, which can be lumped together as 'transaction costs.' Let us go back to our individual, seeking to maximize control of those matters that are important concerns to him. In each matter where he is affected by the actions of others, Dahl suggests, he will wish to form part of a political unit with those others to deal with the matter. In the terms used by economists, this means that the criterion for political units is that they should include all those affected by spillovers generated by some kind of activity. Or, in different terms, the members of the political unit should, between them, internalize all the externalities produced by actions of some kind. Clearly, if an action of a certain kind does not affect others besides the actor, there is no basis for the formation of a political unit to control such actions. This suggests (not surprisingly when one considers the. individualism of the underlying assumptions) that Dahl's principle is the obverse of J. S. Mill's 'simple principle' in On Liberty: the principle that an individual should be free to do what he likes so long

as he does not affect others but if what he does affects others his actions become the legitimate concern of 'society'. In effect, Mill's 'simple principle' is Dahl's principle that externalities should be internalized within the decision-making group applied to the special case where the group internalizing externalities consists of one person only.

Mill himself did not extend his 'simple principle' to say what would be the optimal size of the decision-making unit once the optimal size moves beyond one person. In fact, his discussion of the question of political units in the Considerations on Representative Government gave the heaviest weight, in determining the optimal boundaries, to culture, language and a sense of shared national identity. But it is easy to see how the extension would go.

We have a collection of individuals and the question is what should be the boundaries of decision-making. For matters where each person's decisions affect only himself, each person should constitute a separate decision-making unit. (This is Mill's 'simple principle'.) For matters where any act falling under a given description (e.g. murder, forcible rape, robbery, demanding money with menaces, infliction of bodily harm) is liable to be seriously detrimental to others, everyone would presumably favour a system in which such actions were prohibited by the criminal law and the prohibitionenforced through a police and judicial system. Protection against harmful actions of others may also be provided through the civil law either through injunction (which then makes it a criminal offence - contempt of court - to do the thing enjoined) or through compensation where the harmful action has actually taken place.

In the absence of transaction costs there is no reason why these remedies should be restricted to cases in which the damage is done to

one or few people. The economist E. J. Mishan has suggested that there should be a general principle that the law protects people against the negative 'spillovers' including air and water pollution of all kinds, noise and vibration, visual disamenity, and so on. A situation in which others were not harmed would be the baseline and departures from it would be allowed only if those proposing to inflict the damage were able to offer sufficient compensation to the prospective losers to gain their consent to being adversely affected.

Clearly, it would be necessary for the state to specify in such cases exactly whose permission would have to be obtained by the proposed polluter, which presupposes that they should be a clearly-identifiable set of people. It would be important to the equity and efficiency of the scheme that all those and only those adversely affected should be eligible for compensation.

who will not be adversely affected, this is inequitable and/or inefficient. It is inequitable to the extent that it enables some who will not be harmed to extract advantage as the price of consent, and forces someone who will not be harming them to transfer resources to them. It is inefficient to the extent that, by giving a veto on change to unaffected people it makes it more likely that a potentially Pareto-optimal change will be frustrated. (A potentially Pareto-optimal change is one with the property that the total gains are greater than the total losses so that it would in principle be rossible to redistribute the gains so that some people would be better off and none worse off.)

Conversely, it is also (and perhaps more obviously) inequitable and inefficient if the set of those whose permission is required does not include all those who will be adversely affected. It is inequitable because it opens up the possibility that some people will lose as a result of

another's action without any means of redress, while the actor is enabled to injure others with impunity. It is inefficient, in the sense already defined (that of the 'new welfare economics') because it makes it possible for an action to occur that would not be able to meet the test of giving rise to enough total benefits to allow for full compensation to the losers while still leaving somebody ahead.

Two points should be noticed here, since they will be relevant later. The first thing to notice is that, although there is no necessary reason why those affected adversely by an action should be all and only those residents within a certain geographical boundary, the requirement that the losers from an action should be clearly identifiable and not simply on their own say so, either - inevitably entails that what constitutes 'loss' will have to be 'objective' deprivation. In other words, it will have to be the sort of thing that any normal person would find dangerous, unpleasant or distressing, - in fact the sort of thing that results in successful appeals for reductions in rateable value. ('Normal' should be taken as excluding only fairly specific mental or physical abnormalities so as to avoid letting in 'subjective' considerations by the back door. Thus, someone with strong selfdestructive tendencies might welcome the high risk of death arising from a nearby explosives factory, a deaf person would not mind having jets taking off in his vicinity day and night, someone with perverse tastes might enjoy the sights and sounds of an abaltoir at the end of his garden, and so on.)

The second point to notice is that, strictly speaking, there is no collective decision to be taken here: all that happens is that a set of individuals have to give their consent to a proposal. The collectively-binding decision is the prior one that specifies generally what counts as an action affecting others adversely. After that there are simply a number of separate bilateral negotiations. Nevertheless, the process

outlined fully satisfies the requirement that each person should have some control over events that affect him or her. Indeed, it satisfies it far more than does Dahl's own standard, that a citizen should have the 'capacity to participate effectively in decisions he cares about' (page 133) by voting for respresentatives and the like. For that gives him only 1/noî the control over a collective decision where n is the total number of people taking part, whereas the system outlined enables each individual to prevent undesired changes unless he is fully compensated for withdrawing his opposition.

The principle that people should be prevented from injuring one another may seem obvious, but its obviousness has escaped many people (especially economists of a laissez-faire tendency) when applied to the harmful or unpleasant spillover effects of industrial production.

Sufferers from The argument is that, in the absence of transaction costs,/a non-excludable evil like air pollution could club together to pay the factory to install smoke-control machinery. If they were unwilling to do so this would show that it would not be efficient for the pollution to be checked: the aggregate gain from the pollution must be greater than the aggregate loss. 1

¹See, for example, Otto A. Davis and Morton I. Kamien, 'Externalities, Information and Alternative Collective Action' in Robert H. Haverman and Julius Margolis (eds.) <u>Public Expenditure and Policy Analysis</u> (Chicago: Markham, 1970) pages 74-95, at pages 88-90.

The obvious case against this proposal for dealing with nonexcludable negative externalities is that it is extraordinarily
inequitable. Indeed, it is difficult to see any reason why there should

be a criminal law against demanding money with menaces, blackmail, kidnapping and other forms of individual extortion if it is to be legal to demand compensation for not drenching people with sulphuric acid or exposing them to radioactivity so long as this is done on a wholesale basis. In both cases it can be said that if it is worth paying the ransom then the deal is Pareto-optimal and if it is not worth paying the ransom that shows that it was not potentially Pareto-optimal.

It might be said that the difference lies in the fact that in the first kind of case the evil is threatened with the intent of securing compensation whereas in the second case it is merely incidental to the pursuit of profit from production. But there are two replies to this. First, it is surely a bizarre principle that it should be illegal to impose suffering on others if it costs you something to do it (e.g. the good given to the kidnapping victim) but all right so long as it is profitable to impose the suffering. If one has to choose, it would seem more sensible to say that profitable extortion should be prohibited before costly extortion. And second there is nothing to guarantee that a firm will cause only that amount of injury that is a by-product of the method of production that maximizes its profit. If one could count on that, there would be no necessity for laws against demanding money with megaces, blackmail, kidnapping and so on, except inasfar as these were intrinsically enjoyable activities. If, however, we allow that it may be profitable to threaten loss to others in order to extract a return from them, even if it costs something to mount the threat and (if necessary) carry out the sanction, why should not factories gratuitously create injurious conditions even if they cost a little to produce, in order to get a return from the victims for stopping?

If we ask why in practice non-excludable harm is treated less rigorously than specifically-directed harm within legal systems, part of the answer is, ironically, that the existence of transaction

costs makes it much harder to exploit the threat of causing harm in pursuit of gain when the harm is non-excludable. Non-excludability entails that the harm has to be either inflicted on a whole group of people or on nobody. It cannot be adjusted on an individual basis so that those who have paid are let off, and only the rest harmed. But by contrast a gang running a protection racket (which is in every other respect an identical operation) can be selective: they can wreck the premises of those who do not pay while leaving alone those who do. If I am right, it is an exact reversal of the true position to say that, but for transaction costs it would be quite satisfactory for nonexcludable harms to be bought off by the sufferers. Rather, it is only the existence of transaction costs that makes it tolerable to allow (some) non-excludable harms to be inflicted without the consent or the full compensation of those affected. For the existence of transaction costs does at least rule out the gratuitous infliction of harm and keep it down to that which is profitable in itself rather than as a bargaining counter. Admittedly this is not much, but it is something.

To say that allowing non-excludable externalities would be 'inequitable' is, in fact, putting the point too weakly. There would be no security if one could at any time be faced with a choice between suffering or buying off the infliction of poisonous gas, asbestos particles or radioactive emissions. Thus, there is a deep inconsistency in Buchanan and Tullock's The Calculus of Consent between their initial premise that

¹J. M. Buchanan and Gordon Tullock, The Calculus of Consent (Ann Arbor:

some minimum protection of person and property must be guaranteed by the law and their defence of the position that negative externalities such as air pollution should not be prohibited but should be stopped, if at

all, by those affected paying 'compensation'. For the second simply undermines the first.

It seems unnocessary to spend much time discussing whether or not it is efficient to allow the imposition of negative externalities without consent or adequate compensation. There are technical difficulties about raising the money voluntarily for the relief of a non-excludable harm, which are strictly analogous to those that will be discussed below in the more reasonable context of raising money to pay for a non-excludable benefit. To anticipate that discussion, it may be said that the money for buying off the polluting factory (etc.) can be raised voluntarily if the factory promises to stop polluting if and only if all those affected pay a specific amount, assessed for each individually on a basis of the value of the relative benefit derived from the absence of the pollution.

SECTION III

We have so far been looking at the appropriate way of dealing with negative external effects of actions, that is to say effects which are dangerous, harrful, disgusting, etc. to others other than the actor. Now we need to turn to positive external effects of actions, that is to say effects of actions that are beneficial to others than the actor. There is an obvious asymmetry here with negative externalities. The problem about negative externalities is to suppress them or ensure that the losers from them are adequately compensated. But the problem with positive externalities is simply that they may not occur often enough and that those who provide them may not be compensated for providing the benefits. Thus, the whole range of actions with positive external effects that people do simply because they want to do them or out of affection for the beneficiaries present no political problems: they happen anyway and the more of them the better from any point of view.

effects that the actor would not wish to do for their own sakes but which it is nevertheless feasible for him to be induced to do by his being offered a reward for doing it by the prospective beneficiary or beneficiaries. The act of offering a reward is by definition one with positive externalities so what we have here can be analyzed as an exchange of positive externalities. The only condition that has to be fulfilled for this sort of trading of benefits to occur whenever it is officient for it to do so is that it should be feasible for the positive external

effects of the action in question to be directed exclusively to those who have offered a reward for it. There is no need for any collective decision-making here except for a general law of the usual kind providing for the enforcement of contracts. Thus, suppose that you repair my washing machine: I get the benefit and nobody else - nor, of course, is anyone harmed. Thus there are no effects external to the two of us taken together. Moreover, each of us can keep the status quo (I have a broken washing machine, you do nothing) if we choose to. These two conditions together guarantee that if we reach a mutually satisfactory deal it will be efficient, according to the criterion of potential Pareto-optimality, since ex hypothesi we both gain and there are no losers. The two conditions also assure at any rate prima facie equity, since no third party either loses without compensation or benefits without contributing, and we both feel that what we gave up was worth less to us than what we got in return. (It is only prima facie equitable since it is consistent with these conditions that the parties to the deal might have grossly unequally bargaining positions thus enabling one to obtain an unfair proportion of the total surplus generated by the deal.)

It is worth pausing for a moment to see how naturally all this fits in as a development of Mill's 'simple principle'. The Wolfenden Committee on Homosexual Offences and Prostitution(Cmnd 247, 1957) explicitly basing itself on Mill's 'simple principle', formulated the doctrine that nothing done by 'consenting adults' should be of any concern to the criminal law unless there are other people who are liable to be affected adversely by actions of such a kind. Market dealings (whether in washing machine repairs or in sex) are from this point of view a special case of transactions between 'consenting adults', the special feature being that the consent of one party is bought. (The law

of contracts, incidentally, makes the Wolfenden criteria - 'consenting adults' into conditions of validity: the parties must be adults and must not have been coerced. Thus, Mill's simple principle' that individuals may do what they like so long as they do not harm others is extended to say that pairs or larger groups of people may do what they want so long as they do not (by their collective acts) harm others and so long as the members of the group consent.

So far, then, positive externalities present no difficulty. Problems arise, however, where the benefits are 'non-excludable', that is to say where, if the benefit is to be provided at all, it cannot be restricted in its incidence. Thus, for example, suppose that I am deciding whether to maintain a big flower bed in my garden that gives pleasure to passers by but is not visible except from the street, and I conclude that it would require £50 a year to make it worth my while. Now, it may well be that the total amount that those who enjoy the flowers would think it worth paying to avoid losing them would be substantially in excess of £50. But there is no way in which I can set a price to individuals for looking at the flowers because I cannot prevent those who do no pay from looking just the same. I can, of course, put out a box, invite contributions to the cost of upkeep, and say that the flowers will be maintained only if the contributions exceed £50. Even better, I can ask each contributor to put the money in an envelope marked with his or her name and address, and promise to return all the money if less than £50 is collected and to return the proportional excess to each contributor if more than £50 is collected.

We must observe, however, the consequence of non-excludability.

It entails that the calculation facing each passer-by is deciding whether or not (and if so how much) to contribute is not like that where a price can be charged. If season-tickets at fl per head could be sold with the condition that only those who bought one could see the flowers, each

person would simply have to decide whether it was worth £1 to see the flowers. But because of non-excludability there is no direct connection between what a person pays and whether or not he is able to see the flowers. For whether or not the flowers can be seen (by anyone) depends on the total amount offered by the public, and the caly way on individual's contribution makes a difference is by increasing the probability that the total will reach the minimum necessary of £50. Therefore, even if someone would be willing to pay over fl to see the flowers rather than not see them it does not follow that it is in his interest to offer £1 to the owner of the garden. For it is a priori most likely that the total amount offered by the rest of the passers-by will be either less than £49 (in which case the money will be returned and there will be no flowers) or over £50 (in which case that part of the money over £50 will be returned and there will be flowers). only contingency in which as offer of £1 will make the difference between flowers and no flowers is when the total amount effered by others is not less than £49 but less than £50. The chances of this must in general be taken as low, which suggests that it would not be rational (on a basis of self-interest) to offer a contribution.

Since all the passers-by who enjoy the flowers are subject to the same calculation, the total amount offered is likely to be far less than the total amount they would between them be willing to pay.

Thus, even if they would be willing to pay £50 they are unlikely to offer it. As a result they will suffer collective frustration.

This kind of incentive-structure, in which each person, by pursuing his own advantage, finishes up with an outcome he likes less

than one attainable with a different incentive-structure, is an n-person prisoners' dilemma. (See the paper by Professor Boudon for a numerical illustration). The solution, therefore, is a change in the incentive-structure so as to enforce the payment of a total of £50 by the passers-by. Ex hypothesi there is some arrangement of taxes to be levied among them that they would all prefer to the alternative of no payment and no flowers.

Nothing except a tax, levied proportionally to the benefit arived by each person, and designed to raise a total of £50, can ensure that both efficiency and equity are satisfied. In particular, if voluntary contributions are relied on, the most likely result is that the requisite £50 is not raised. This is inefficient in the sense of 'efficiency' (potential Pareto optimality) that we have been using, because those who gain from the existence of the flowers could compensate the owner of the garden for growing them and still be ahead. In the more unlikely event that a sum of £50 is raised voluntarily, there is no reason to expect that this will have been contributed in an ϵ uitable way, that is to say paid for by the beneficiaries in the ratio of their benefits. It is much more likely to be brought about by a small number of those who feel strongly about the advantage (either as large beneficiaries or nartially out of altruism) agreeing together to pay the money, thus enabling the rest of the passers-by who enjoy the flowers to do so without contributing to the cost.

An ordinary tax, however, raises money not by offering a benefit in return for payment but by offering the absence of sanctions. In other words, the proposition presented to the taxpayer is not 'If you don't pay you won't get such-and-such publicly-provided benefits' but 'If you don't pay you'll be jailed'. There is therefore nothing built into the structure of ordinary taxation that requires the benefits each

person expects to derive to exceed the cost to him of contributing to it. Is there, in the absence of transaction costs, a possible way of ensuring that the connection does hold, while at the same time avoiding the perverse effects of calling for voluntary contributions in the way already discussed?

I think the answer is that we can find a way if we reflect on the way negative externalities were treated, and look for an analogy. It will be recalled that the factory wishing to pollute the local air had to buy permission from the local residents by offering adequate compensation to each. In contrast, an analogue of the 'voluntary contribution' method we have considered for raising the cost of the garden would be for the firm to say how much in total it would pay and for each resident to send in his demand for compensation. The analogous rule of for proceeding would then be that, if the total claims amounted to less than the sum offered the payments be scaled up proportionally; but, if the sum of the claims exceeded the amount offered, there would be no deal and the factory would not be built. Under these circumstances each person would have a strong incentive to pitch his demand for higher than the minimum needed for compensation. For, by parity of reasoning to that exhibited in the garden case, each person had to calculate the probability that a demand of given size will make the difference between the total amount demanded falling below and falling above the total amount offered - and this probability is likely to be low. The likelihood is therefore that, under such a scheme, the factory would not get built even if the factory could afford a payment to each prospective sufferer that would make him better off than the status quo.

The essence of the scheme actually proposed - that the factory

makes an offer to each affected householder - is that each householder

has a veto over the operation of the plant, it may appear that the firm is in a weak bargaining position and so it is relatively to the consequences some alternative rules. But, so long as the amount it offers each resident really is enough to make it advantageous on balance for each to accept, weakness is strength since the firm can put each person on the spot and say 'If you want the money you have to accept these terms.' The equivalent in the garden case is that the person with the garden has to take the initiative, find out who the passers-by are who enjoy the garden, and send each of them a demand for a specific amount of money (based on an assessment of the value of the benefit derived from the garden) with a statement to the effect that the garden would not be maintained unless everyone contracted to pay the amount demanded. Provided the demands were correctly calculated in every case, this would give everyone a sufficient reason to pay up. But the proviso is crucial, because even one demand pitched higher than the value of the garden to the purson concerned would result in the collapse of the scheme. (It would not, however, matter from the efficiency viewpoint if some were underassessed or even missed out altogether if the rest could be changed enough to make the total sum high enough. But it would be a departure from equity, of course, in that there would be some free (or cheap) riders, enjoying benefits for which they had not paid their share.)

In principle, then, it is possible to deal with positive externalities so as to satisfy the requirements of efficiency and equity, as we have defined them so far. Efficiency demands that where the gains to others arising from an action could more than compensate the actor for providing the benefit it ought to be done. Equity demands that if a benefit is derived by another from an action that other should provide recompense to the actor. Where the benefits have the property of excludability, the ordinary mechanism of exchange among individuals

is available: each party to the exchange decides if the benefit to him is worth the cost to him; if it is he agrees to the exchange and if both agree they have a deal which is efficient and at least prima facie equitable. With non-excludable goods, as we have seen, this simple solution breaks down but it is in principle possible to collect contributions from beneficiaries by making each one face a choice between paying a set amount and not getting the benefit - though, of course, Gobody else then gets the benefit either. If the non-excludable benefit is provided on these terms, it must be efficient and at any rate partially equitable. (It need not be fully equitable since there may be undercharged or uncharge beneficiaries). But, of course, if the charges demanded are incorrectly assessed a non-excludable benefit that it would be patentially efficient to provide may not be provided.

It should be clear that, if transaction costs could be discussed. this analysis would provide a basis for a voluntary system of taxation, (in a special sense of 'voluntary') to pay for the state to provide nonexcludable benefits for the population either directly by paying for its employees to act in ways productive of non-excludable benefits or indirectly by giving subsidies to private individuals or corporations to carry out functions giving rise to non-excludable benefits. the present analysis, the two would not be sharply distinguishable.) The relevant sense of 'voluntary' is not, it should by now be apparent, the ordinary one whereby individuals give as much or as little as they like (whether anonymously or publicly makes no difference at this point) and what services are provided depends on the total amount given. Rather it is the sense in which contributions to the garden were voluntary in our final procedure. That is to say, the state would propose to provide a certain benefit provided all the citizens were prepared to pay the tax assessed against them in respect of this benefit. 1

¹See Knut Wicksell 'A New Principle of Just Taxation', in R. A. Musgrave and A. T. Peacock (eds.) Classics in the Theory of Public Finance (London, 1958) pp. 72-118.

The tax for some people might well be zero for benefits with a local or specialized inpact, but it would obviously be crucial that those preparing the demands were able to estimate correctly (or at any rate not overestimate) what each benefit would be vorth to each person.

Again, one overestimate and the benefit in question would have to be withdrawn from everybody.

It is, however, it portant not to get carried away by the fact that we have found a role for the state within the present framework. It may rashly be supposed that the rationals for the activities of real-life states is to be found in non-excludable benefits. Such a view can not withstand examination. Almost all things supplied free or at subsidized prices by modern states are technically 'excludable', and the decision to make them into 'public goods', financed by taxation rather than by sale to those who are prepared to pay and only then, is a political decision, not a technical necessity, and as a political decision we shall have to look for its foundations elsewhere.

SECTION IV

So far I have been asking what would be the appropriate political institutions for a society whose members were concerned (1) to protect themselves against one another and (2) to ensure that acts beneficial to others would be carried out provided that the benefits exceeded the costs could. And I have argued that, in the absence of transaction costs, there would be something close to unanimous agreement on a scheme with these four features: (1) laws prohibiting acts of a kind detrimental to others, enforced out of taxation levied on those in the area covered, (2) a provision whereby the constraint could be removed if all those affected agreed to accept compensation for the damage done then, (3) no limitations on what two or more individuals could do together (whether or not economic reward entered in) so long as all those concerned consented and nobody else was adversely affected, and (4) provision for voluntary taxation, of the special kind described in section III, to pay for non-excludable external benefits of actions, whether supplied privately or by the state.

It is clear from this that the question of the boundaries of political units would not, in such a society, be of any great significance. The laws would be the same everywhere, and the requirement of unanimous consent for inflicting damage or raising contributions to the costs of non-excludable goods would, in effect, create a unique ad hoc decision-making body for each issue.

How are things changed when we drop the stipulation that there are no transaction costs? I think that we now modulate into something like Dahl's prescription as far as the areas of decision-making units are concerned, but not necessarily into the ideal of decision-making within those units in accordance with majority preferences for outcomes.

If we look at the supply of non-excludable benefits, we see that the introduction of transaction costs moves some things out (so that nothing gets done about them) and other things into the category out of that of excludable (marketable) benefits.

Thus the example of the flowers in the garden that I developed in Section III is in fact a typical case of a positive externality which public authorities do not normally intervene to secure. In some places it is a criminal offence to allow one's garden to become infested with weeds, because they are liable to propagate in surrounding areas. Public intervention is thus normally restricted to preventing harm to others and only rarely extends beyond, as in Canberra, where the government presents each buyer of a new house with a tree, as a contribution to the city's amenities. Similarly, public authorities may intervene to prevent the construction of an eyesore or the demolition of an aesthetically or historically valuable huilding, but seldom offer inducements to owners to keep their buildings in a good state of external repair and decoration, though again exceptions may be found.

At the same time as many non-excludable benefits drop out of consideration, benefits that are in principle excludable but expensive or inconvenient to sell on the market come in. A number of the usual public services (especially parks and roads) fit this description, as do subsidies to operation with a 'stand-by' value like the railways, But if these are to be 'rublic goods', supplied free of charge (or below cost price) to all corners indiscriminately, how are they to be paid for? It is here that the boundaries of political units become important. We must, I think, accept that the existence of transaction costs rules out the nice adjustment of demands for payment to take account of the benefit each person derives from the free or cheap . service. Taxes therefore have to be levied on some objective basis equal payments, payments related to income, payments related to the ownership of property, or indirect taxes like sales taxes or VAT. This entails, as we have seen, that decisions to raise money cannot depend on unanimous consent. For, however carefully the constituency of voters is drawn up, the basis cannot be that only those are included who will benefit by more than the amount of tax it is proposed they shall pay. This would reintroduce transaction costs by an alternative route - and it should be noticed that when we speak of 'transaction costs' here we are often referring to operations that have an infinite cost, in other words operations that cannot be carried out at any cost. At the same time, however, equity and efficiency both require (1) that the area to which the tax is to be levied should correspond as far as possible with

'principle of fiscal equivalence) and (2) that the requirement for making the expenditure should be as near as possible to unanimous consent.

Take first the size of units. It is inequitable if the beneficiaries include some who did not contribute to the cost (the political unit is 'too small') and also if some who contribute to the cost cannot enjoy the benefit (the political unit is 'too big'). What about efficiency? If the political unit is 'too small' benefits may not get provided, because there is no way of taxing all the beneficiaries, even though the total benefits would outweigh the total costs. If the political unit is 'too big', benefits may again not get provided even if the beneficiaries would be willing to pay for them, because those who do not stand to benefit but are liable to be taxed to pay for then will vote against.

The result of departing from unanimity is to open up possibilities of both inequity and inefficiency, because those who stand to benefit from a piece of public expenditure can raise part of the costs from non-beneficiaries, and it may be that the total gains of those who will benefit do not exceed the total cost that has to be raised in taxation.

It may appear on that the two forces - heterogeneity of interests within the political unit and departure from decision-making by unanimity - tend to offset one another. But such a view would be superficial.

It is quite true that, for any given size of majority required for a collective decision, public expenditures will be lowered by heterogeneity within the political unit. And it is also quite true that, for any given degree of heterogeneity within the political unit, public expenditures will be raised by lowering the size of majority required for a collective decision. But the two forces could only be said to

cancel one another out in any but this gross way if the extra public expenditure that would be generated by lowering the majority required for a collective decision were the same public expenditure that would be approved in more homogeneous units by bigger majorities. As far as I can see there is no reason to expect this to be so. The two forces might very well instead reinforce one another: thus, a majority in one part of the political unit could consistently vote for measures whose benefits were concentrated in their area. (The Royal Borough of Kensington and Chelsea is perhaps an illustration of this.) In fact, there is a lot to be said for the proposition that the more homogeneous the political unit is, the less it matters what majority is required for a collective decision.

SECTION V

Much the same analysis can be applied to negative externalities.

Pirst, because of transaction costs (litigation, political action, etc.)

it seems unavoidable that some negative externalities have to be permitted without compensation. 'De minimis non curat lex.' And, as a result the quality of life in a society will demand on people behaving better than the minimum the law demands even more than it would if all that were at stake were simply a matter of a lack of inducements to provide positive externalities. Second, the provisions for compensation, where the negative externality is non-excludable, will need to be modified to take account of transaction costs. Professor Mishan, who proposed the principle that consent should be required from all those suffering from a negative externality, urged that no concessions should be made to allow transaction costs: the person proposing to inflict the damage

should pay the full transaction costs of obtaining unanimous consent from those liable to be affected. There is, it seems to me, nothing inequitable about this but it does open up prospects of extreme inefficiency. For the transaction costs of obtaining consent for a plant affecting air, water, visual amenity and so on over a wide area would be enormous and perhaps infinite. The result would therefore be that a great many things that would satisfy the criterion of potential Pareto-optimality (the gainers could more than compensate the losers) will fail to get done.

It seems fairly clear that Mishan himself would not be worried about this, but if he wishes to take the position that technology should be restrained more that the people in a country want (taking full account of the impact of spillovers) this needs arguing on overtly elitist grounds - and Mishan has always denied that he makes use of eligist premises.

Here is the point at which we should take up the analogy with the proposed treatment of publicly provided or publicly subsidized services. The idea there was that the area of the political decision—making unit should as far as possible coincide with the area containing those who stand to benefit from the service and that some majority greater them a simple majority should be required to consent before the taxation can be validly imposed to pay for it. The analogy is that the area of the political decision—making unit empowered to decide whether or not a substantial amount of damage or disamenity is to be allowed should coincide

as closely as possible with the area in which the damage or disamenity will be suffered, and that any proposal to permit its infliction should need some majority greater than a simple majority.

Such a scheme would keep transaction costs within reasonable bounds (compared with a requirement of unanimous consent) but at the same time it would provide a good chance for both efficiency and equity to be satisfied. It should be said, however, that these desirable effects require that the political unit should have the authority to extract some quid pro quo for making an exception to the ban on negative externalities. If it could do nothing except say 'yes' or 'no' then this would, of course, not be at all efficient since all proposals for inflicting negative externalities would presumably be turned down by those adversely affected, whether or not the criterion of potential Pareto-optimality was satisfied. I assume, therefore, as part of the scheme, that the permission to cause damage or disamenity can be sold. as a once-for-all payment for a licence or as a continuing tax to be paid to the political unit. Under these conditions, we can say that any proposal that gets accepted must be reasonably efficient, since we know that the gainers were willing to compensate the great majority of the losers, and reasonably equitable since most of the losers do in fact obtain componsation that they regard as adequate.

As with the supply of public services, we can show that deviations from the 'right' boundaries for the political unit open up increased possibilities inequity and inefficiency. If the unit is 'too small' it will be possible for permission to be obtained for imposing a negative externality by buying the consent of only some of those who will be adversely affected, which means that the gainers may not be able to compensate the losers and some of the losers will not in fact be compensated. If the unit is 'too big', then if the rate of compensation

is the same for all those in the unit, some potentially Pareto-optimal changes will be vetoed, because the gainers cannot afford to compensate the losers adequately to gain their consent and also offer the same rate to the 'free riders' on the compensation bandwagon, who are not affected adversely but are entitled to share in the proceeds of the tax all the same. If the proposers of the externality can afford to pay those who are in the unit but not affected at the same rate as those who are, this means that an efficient change takes place but is still inequitable in that it provides a windfall for those who were unaffected.

clearly, the inefficiency that we predicted from a 'too large' unit - that potentially Pareto-optimality changes may be stopped - occurs only if those who are really affected adversely by the proposed change form a large enough proportion of the unit to be able to block unsatisfactory terms for compensation. For example, if they form only 20% and a 75% affirmative vote is required, they may regard the compensation as insufficient but the unaffected 80%, who stand to get something for nothing, will presumably vote in favour of accepting the compensation.

It may therefore, again, be thought that the answer to heterogeneity is to lower the size of majority required. But there are two objections to doing that. First, it adds a new inequity to the existing one: those who don't suffer continue to get paid but those who do suffer now lose any guarantee that the compensation they receive will be adequate. In effect, the majority are being bribed to sell out the minority who actually stand to lose.

This points clearly to the second objection: there is nothing in such a state of affairs tending to ensure that only efficient

(potentially Pareto-optimal) changes go through. If the majority of those unaffected by a proposal negative externality is larger than the majority required to constitute an acceptance of the compensation proposed for the negative externality, this leaves it open for even grossly inefficient proposals to pass. Those who unaffected may be willing to accept a level of payment for consent which is much lower in aggregate than the net loss (loss minus compensation) suffered in aggregate by those who are affected.

An example is provided by Jane Jacobs in her attack on the ideals and practice of contemporary city planning, The Death and Life of Great American Cities, (Harmondsworth: Penguin, 1965). Areas . considered ripe for (in British terminology) 'comprehensive redevelopment', that is to say demolition and replacement by 'projects', are acquired compulsorily under the power of eminent domain, but the public authority has to pay 'only for what it acquires for itself and not for what it takes from the owner' (page 326). If it acquires a business, it pays for the building but not the loss of business or goodwill, and this may be only a fraction of what it costs to buy the business as a going gencern. She concludes that these projects' are inherently wasteful ways of rebuilding cities, and in comparison with their full costs make pathetical contributions to city values (and thus increased tax returns). At present, society is protected from these facts of life because so high a proportion of the costs is visited upon involuntary victims and not officially added in. But the cost is there.' (page 327) Thus, her charge is that the process is inefficient and inequitable. It is inefficient in that the gainers would not be prepared to compensate the losers: 'Ware the involuntary subsidies which make these schemes possible included in public costs, the enlarged public costs would bear no

conceivable relation to anticipated tax returns' and the subsidies required on public housing would be 'politically unrealistic' (page 327). And it is inequitable because 'the community as a whole should bear the expense of community progress and that cost should not be imposed upon the unfortunate victim of community progress' (page 326).

This is not exactly like the cases considered in the text, in that the compensation is offered selectively, but the essential point is the same: the supposed benefit (in increased taxes and the 'greater good' of the city) is widely diffused but the losers are concentrated and amount only to a small minority of the members of the political unit.

The pp blem of 'too big' units is not, therefore, to be solved by removing the safeguard that those who are going to be damaged should themselves be prepared to accept the proposed Ledel of compensation before the imposition of the externality can be allowed. But it can be alleviated if we drop the assumption that whatever compensation is paid must be paid in a way that ignores the fact that some members of the political unit may be deeply affected, some less, and some not at all. Obviously, this is the easiest practice first because it saves the need to make discriminations and second because (if the political unit provides public services) the taxes received can simply go into the general treasury to offset the costs of providing public services. But in spite of increased 'transaction costs' it may be quite feasible to related proposed compensation in a rough and ready way to severity of impact from, say, aircraft noise or air pollution. If it is, then both efficiency and equity can be satisfied fairly well even in units that are 'too big': so long as the estimates of the different levels of compensation required are about right, potentially Pareto-optimal

measures should go through, and compensation will actually be related to loss.

Thus, we can get a solution which maintains some Teatures of the scheme for a world without transaction costs, where compansation offers were tailor-made for each individual affected, but adopts it to take account of transaction costs. It may have occurred to the reader that the same kind of solution is also available for gaining assent to paying the cost of supplying public services in areas that are 'too big'. However, there is a difference that make such an approach less attractive for public services. It seems to be a characteristic of human beings that their dislikes are more uniform than their likes. It is, I argued in Section II, possible to talk about things that any 'normal' person would prefer not to be subjected to. It is much harder to find a list of public services (except the maintenance of law and order, which is really aimed at preventing acts with negative externalities anyway) that any 'normal' person '-uld wish to have provided. Moreover, where the impact of negative externalities often depends on location, the impact of positive externalities depends much more on taste. Everyone has to breathe so the main determinant of impact is location. But how highly you value a free library, park, winming pool, etc., depends partly on how near it is but more decisively on whether you would want to use it even if it were next door. The problem of raising the money for public services is therefore inherently less tractable because the criteria for benefit are less 'ojbective' than the criteria for damage or disamenity. But where, for example, location is a significant determining factor it may be that there is room for a variation in tax rates within the political unit.

At the same time, however, it may be observed that this source of heterogeneity also lends itself most easily to treatment by operating on

the supply side. If most of those in the unit have to agree to raise the taxes for a localized amenity, the only way of getting it through may be to split the location of the amenity so as to spread the benefit or (if there are big economies of scale) put together a package in which one area gets one amenity centred in it and another area a different one. Thus, the need for differential taxes is smaller (where they could be applied) than the need for differential compensation.

The upshot of the discussion so far, then, is to suggest that the logical complement of Dahl's criterion for the boundaries of political units is not democracy but father that something approaching unanimous consent should be required for public expenditures and for deviations from laws prohibiting the infliction of harm on others. I have not so far dealt with Dahl's own modification of his criterion for the size of units, that for practical reasons concerning communication and control there should be only a small number of multi-function political units. The point is clearly well taken, and it is a relief to observe, therefore, that the most serious problems of units that are 'too big' can be greatly alleviated if there can be differential compensation or (somewhat less promising) differential taxes. At the same time I think that Dahl underestimates the extent to which the multifunction authorities might create workable ad hoc political units.

At the same time, I am sure that Dahl, with a couple of vague references to problems of 'loyalty' to political units, is right to believe that ad hoc units, their boundaries determined by the scope of non-excludable positive and negative externalities, have a limited usefulness in dealing with real political problems. However, this is an instance (not the first) in which Dahl's own sense of reality has run ahead of his theoretical apparatus. In the rest of this paper, therefore, I want to ask what are the deficiencies of the theory of political decision-making so far developed, and what implications for the boundaries and decision-rules of political units are generated by a more comprehensive theory.

Perhaps the best place to start is to notice how closely the general lines of the theory developed in this paper coincide with those of the social contract theorists of the seventeenth and eighteenth

centuries. In particular, we may say that (in common with almost all American political thinking from pre-independence times until today) it is an adaptation of Locke's Second Treatise of Government. Indeed, I think we can use that analysis in the earlier part of the paper to throw light on a long and inconclusive debate: did Locke espouse the 'doctrine of majority rule' or not? The reason for the inconclusiveness is, I suggest, that there is an inconsistency between Locke's premises and his conclusions. All Locke's premises lead to a case for political decisions to be taken by unanimous agreement or (to allow for transaction costs) some approximation to it. He postulates a world of individuals concerned above all with protecting themselves against one another and on the basis of this he argues that everyone would agree to having laws enforced to protect everyone's person and property. He also says that taxation (to pay for public goods, including of course the cost of law and order' and defence) should be raised only with the consent of all the taxpayers. But he then waters it down, first to the representatives of tax-payers and then to a majority of representatives of tax-payers. His arrival at this position from the requirement of unanimity suggested by his premises is prepared for only by the casual remark that since one cannot expect unanimity the only alternative is decision by simple majority. It is certainly true that, once a retreat is made from unanimity there is nowhere short of a simple majority that has the same quality of 'obviousness'. Nevertheless, he makes it quite explicit that any departure from unanimity is an unfortunate concession to the difficulties of reaching agreement, so decision by simple majority has no special significance except its 'obviousness'. Thus, Buchanan and Tullock, in their Lockean exercise The Calculus of Consent, seem to me to be pursuing the logic of Locke's premises more consistently

than Locke himself did, when they write: 'The individualistic theory of the constitution that we have been able to develop assigns a central role to a single decision-making rule - that of general consensus or unanimity. The other possible rules for choice-making are introduced as variants from the unanimity rule. These variants will be chosen, not because they will produce "better" collective decisions (they will not), but rather because, or balance, the sheer weight of the costs involved in reaching decisions unanimously dictates some departure from the "ideal" rule... At best, majority rule should be viewed as one among many practical expedients made necessary by the costs of securing widespread agreement on political issues when individual and group interests diverge' (page 96).

1

My reason for bringing in Locke here is that, unlike most of those who have followed him (whether they were aware of it or not) he recognised very clearly what had to be established if a political theory of the kind he was putting forward in the Second Treatise was to stand up. And it is a sign of his practical intentions that, having seen what foundations were needed, he set out to provide them. These two requirements are (1) that people should regard as politically relevant only what affects them personally and (2) that people should not regard the question of the distribution of property or income as politically relevant.

Locke addressed himself to the first of these requirements in his retters on Toleration. At the time when he was writing, religion provided by far the biggest risk that people would concern themselves with the behaviour of their neighbours on the basis of a conception of the way the society should be rather than on the basis of regarding other people as sources of privately-consumed positive and negative externalities. He therefore set himself the task of arguing that religious

observance should be assimilated to the same individualistic criteria for political intervention as would be applied to, say, blocked-up drains. That is to say, unless you are actually harmed (in the kind of straightforward sense we have been using so far) by someone's religious practices, you have no grounds for trying to stop them. You could, for example, object to being kept awake by their chanting or to having to smell their burnt offerings, but not to the content of their beliefs or observances as such.

The attempt to rule out the distribution of property or income as possible subjects Cor collective decision-making was carried out in the opening sections of the Second Treatise itself. The object of Locke's argument here is to establish that 'out of the bounds of society' men could have property, first in what they had acquired by their own labour and then, by 'consenting to the use of money', in land itself and in the store of generalized value represented by money. Men therefore come into political society the better to secure their (pre-existing) property. The state does not create property but underwrites it. The significant conclusion which Locke wants to get out of this confusing and implausible rigmarole in that distribution is not, in any existing society, an issue that can legitimately be raised politically. Taxes, as we have seen, need to be levied to pay for the operating expenses of government - mainly 'law and order' and defence - but they would ideally be levied only with the unanimous consent of the taxpayers and are certainly not supposed to be used to raise money for redistribution of property or income.

For almost three centuries now liberal thinkers have been trying to improve on Locke's arguments in favour of these two fundamental propositions, and I think the net result of their labours has been to add a little to his arguments on the first and subtract from his

arguments on the second without, however, finding any very satisfactory replacement. More specifically, the argument in favour of religious toleration has been refined and generalised without departing from the main lines of Locke's treatment. In particular, I should make so bold as to suggest that nobody (including Rawls in his massive (Theory of Justice) has succeeded in explaining why one should, for the purpose of politics, be exclusively concerned with one's own private condition and should not have a view about the kind of society in which one wishes to live, if one does not happen to be predisposed to be interested in one's society only insofar as it impinges directly upon oneself. On distribution, conservatively-inclined liberals from David Hume to Bertrand de Jouvenal have liked Locke's conclusion but, being embarrassed by the maxifest implausibility of the deduction of property from agreements in the 'state of nature', have fallen back on the negative argument that no rational basis can be given for holding any distribution to be just or unjust, so in the interests of stability the only thing to do is that everyone should accept the status quo, whatever it may he.

Now, it is my contention in this paper that any realistic theory of politics must come to terms with both of the phenomena that Locke sought to rule out. It is legitimate to think that the world would be a better place if everyone were to 'cultivate his own garden' and pay no attention to social issues (except inasfar as they affected him personally) or to distribution. But it is, I suggest, not sensible to design politica! institutions on the assumption that other people will follow your advice when it is manifest that they will not.

Thus, Buchanan and Tullock, in <u>The Calculus of Consent</u>, after a brief (and inconclusive) discussion of the problem posed by people wanting to legislate on a basis of what they call 'morality', say

'Normally, of course, there is sufficient standardization of moral values over the population of a community (read: political unit) to prevent serious issues of the sort posed from arising' (page 270).

Anyone who can believe that can believe anything.

As soon as we adopt a perspective in which these phenomena are recognized as important (and, as far as we can see, enduring) features of the real world, we have to regard as extremely peripheral the kind of considerations we have so far applied to the question of optimum boundaries and decision-rules for political units. The main determinants of political boundaries and rules of decision-making, as they are, and as they ought to be, are precisely the two phenomena excluded by Locke. It is these two phenomena that are the stuff out of which history is made. Revolutions, civil wars, international wars, wars of national independence - these come about not because negative and positive externalities have not been dealt with in the optimal way (however much noise outraged ratepayers may make) but because people have passionately opposed ideas about the sort of society (or the sort of world) they want to live in and about the justice or injustice of alternative distributions of property and income.

The only effect of slighting these forces is that one produces constitutions that are like houses of cards. The history of British postwar decolonialization, for example, is littered with the debris of constitutions that were based on exactly the same premises as Dahl uses in Size and Democracy. The civil servants and their academic advisors assumed that wherever there was an interdependence there should be a political unit, and imposed federations. At the best these were scrapped by mutual consent (Caribbean, Rhodesian) but their full potentiality for mischief was exhibited by the bloody civil war in Nigeria.

I do not wish to deny, let me make it clear, that there are real advantages in getting right the rules for dealing with actions of a kind that have effects that are regarded by almost everyone as harmful or as beneficial. And I think that a society would be improved if greater opportunities were provided, along the lines subgested, for local initiative in stopping unwanted changes and raising the money for desired change. But I do wish to insist that for most societies the handling of these questions must be subordinated to the task of regulating conflicts between communities and/or over distribution. In the course of regulating these conflicts positive and negative externalities of the kind we have been considering may get dealt with in ways that economists will regard as hopelessly inefficient. But if these arrangements are part of a package that maintains social peace without repression, they may be cheap at the price.

Thus, we may see industrialization projects shared out between the regions of a country even though the contribution to GNP would be much greater if all the investment were concentrated in one area where the return on capital is highest. We may see all kinds of wasteful duplication of public facilities - wasteful, that is, if we assume that the only object(is to provide a generic service (say, 'education') and ignore the fact that plural provision may be the price of any modus vivendi between religious or linguistic communities.

Fortunate is the society that can afford the luxury of focusing its politics around the question of the optimal treatment of such externalities. Fortunate too would be a world in which communal and distributive issues had been defused to the point at which the problems discussed earlier had emerged as the politically central issues. But the world we actually inhabit is very far from being such a world.

Dahl is right, in <u>Size and Democracy</u> to draw attention to the existence of many world-wide interdependencies: 'the diffusion of nuclear weapons,

nuclear explosions, air and water pollution, security from aggression, international trade, gold supply, monetary stability, foreign investment, international travel, exploitation of seas, oceans, and continental shelfs, and so on.' (page 129). Unfortunately, the existence of all these grave negative externalities that countries can impose on one another does not necessarily lead by Dahl's answer of a world-wide political unit to deal with them, because there are also communal and distributive conflicts at work.

It is my contention that the Lockean position rests either on much too narrow conception of the range of politically important desires or on the view that there are relatively compelling arguments to show that those whose desires cannot be accommodated within the Lockean framework should refrain from pressing for their satisfaction by political means. Unfortunately, I have already imposed too much on the patience of my readers, so I must leave for another occasion a discussion of the boundaries and methods of decision—making indicated for the resolution of conflicts between such desires.