CHAPTER 9

Exit, voice, and disloyalty

> Among the laws that rule human societies there is one which seems to be more precise and clear than all others. If men are to remain civilized or to become so, the art of associating together must grow and improve in the same ratio in which the equality of conditions is increased.
>
> Alexis de Tocqueville

In his book *Exit, Voice, and Loyalty* (1970), Albert Hirschman developed the useful distinction between processes in which individuals express their preferences via entry or exit decisions, and those in which some form of written, verbal, or voice communication is employed. An example of the first would be a market for a private good in which buyers indicate their attitudes toward the price-quality characteristics of a good by increasing or decreasing (entry or exit) their purchases. An example of the exercise of voice to influence a price-cost nexus would be a complaint or commendation of the product delivered to the manufacturer. A necessary condition for the effective use of exit is obviously that the potential users of this option be mobile: and full mobility of both buyers and sellers (free entry and exit) is an assumption underlying all demonstrations of market efficiency. In contrast, the literature focusing on voting processes, public choice and political science, has almost exclusively assumed (most often implicitly) that exit is not an option. The boundaries of the polity are predefined and inclusive; the citizenry is fixed. A citizen is at most allowed to abstain from participating in the political process, but he cannot leave the polity to avoid the consequences of its decisions.

Given the assumption of fixed boundaries and citizenry, the characteristics of a pure public good, nonexcludability and jointness of supply, require that a collective *voice* or nonmarket decision process be used to reveal individual preferences and achieve Pareto efficiency, as Samuelson (1954) emphasized. But many goods are "pure" public goods in a limited sense only. For these goods, the nonexclusion principle and/or the jointness of supply property may not be applicable over the full range of possible distribution and production alternatives. For these quasi- or local public goods, the possibility may exist for employing *exit* as an alternative or complement to the *voice* process. These possibilities are reviewed in the present chapter.

9.1 The theory of clubs

Consider the effect of retaining only the joint supply property of public goods. Exclusion is possible, but the addition of a new member lowers the average cost of the good to all other members; that is, there are economies of scale. If average costs fall indefinitely, the optimal size of the consumption group is the entire population, and the traditional public good problem exists. If they eventually stop falling or rise, either because scale economies are exhausted or because of the additional costs of crowding, the optimal size of the consumption group may be smaller than the population. When those who do not contribute to the costs of providing the public good can be excluded from its consumption, the potential exists for a group of individuals to agree voluntarily to provide the public good only to themselves. We shall define such a voluntary association established to provide excludable public goods as a *club*. Although we shall generally assume that the provision of the public good to club members involves at least some fixed costs, and perhaps some falling variable costs, it should be noted that the public good provided by some *social* clubs consists entirely of the presence of the other members of the club. A bridge club is an example. Here there may be no costs, other than time, to providing the public good, and no benefits other than those arising from the association with the other bridge-playing members. But exclusion is possible, and the analysis of these clubs parallels that of the more general case of interest here. Voluntary associations to provide (or to influence the provision of) nonexcludable public goods do not meet the definition of a club employed here, although these associations sometimes call themselves clubs (for example, the Sierra Club). These associations typically attempt to influence the provision of the public good by some other body, such as a state or national legislature, and are treated here as interest groups rather than clubs (see Chapters 15, 20, and 21).

Buchanan (1965a) was the first to explore the efficiency properties of voluntary clubs using a model in which individuals have identical tastes for both public and private goods. To see what is involved, consider the example Buchanan first employed, the formation of a swimming club. Assume first that the size of the pool, and thus its total cost (F), is fixed and the only issue to be decided is the size of the club. Figure 9.1 depicts the marginal benefits and marginal costs from an additional member as seen by any other member. Given identical tastes and incomes, it is reasonable to assume equal sharing of the costs. The marginal benefit to the first member from adding the second member to the club is the saving of one half the cost of the pool, that is, $MB = F/2$. The marginal benefit of a third member to the first two is the additional saving of one third of the cost of the pool ($F/3$). The additional benefits from adding new members, the savings to the other members from further spreading the fixed costs, continue to fall as the club size (N) increases, as depicted by MB in Figure 9.1. The marginal costs of a new member are given by MC. These are psychic costs. If individuals prefer to swim alone, these will be positive over the entire range. If individuals enjoy the company of others in small-enough numbers, the marginal costs of additional members will be negative over an initial range of

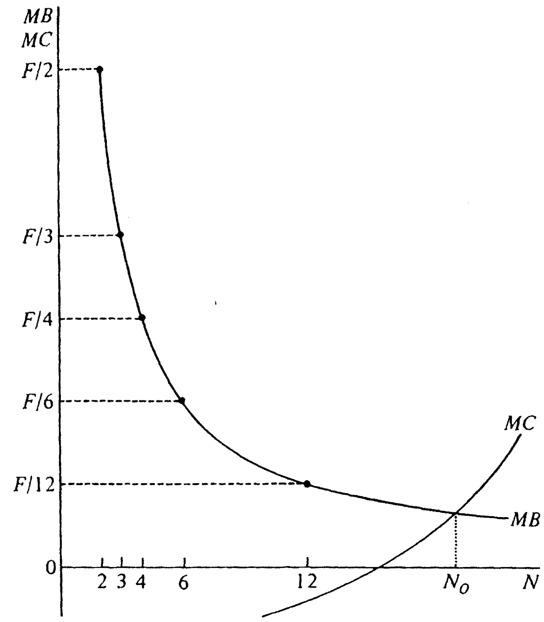

Figure 9.1. Determination of optimal club size.

club sizes. Eventually, the positive costs of crowding will dominate, however, and the optimal club size, N_0, is determined where the marginal cost of an additional member from enhanced crowding just equals the reduction in the other members' dues from spreading the fixed costs over one more club member.[1]

Figure 9.1 can also be used to depict the polar cases of pure private and pure public goods. For a pure public good, the addition of one more member to the club never detracts from the enjoyment of the benefits of club membership to the other members. The marginal cost schedule is zero everywhere and coincides with the horizontal axis. The optimal club size is infinity. For a pure private good, say, an apple, crowding begins to take place on the first unit. If a consumer experiences any consumer surplus from the apple, the foregone utility from giving up half of his apple exceeds the gains from sharing its costs and optimal club size is one. Even with such seemingly private goods as apples, however, cooperative consumption may be optimal. If, for example, the unit price of apples is lower when sold by the bushel, the distribution of apples exhibits joint supply characteristics and might dictate optimal-sized buying clubs of more than one.

The theory of clubs can be extended to take into account the choice of quantity and other characteristics of the collective consumption good. This extension is, perhaps, most easily undertaken algebraically. Let a representative individual's utility be defined over private good X, public good G, and club size N, $U = U(X, G, N)$. Let the cost of providing the public good to the club include a fixed cost, F, and a unit cost (price) of P_g. Assume that each individual has not only the same utility function U, but the same income Y, and that each pays the same fee, t, for membership in the club. In deciding what level of public good to provide and what size of club to

[1] See McGuire (1972, pp. 94–7) and Fisch (1975).

establish, we assume that the utility of a representative club member is maximized. This objective might arise as the conscious choice of the founding club members, or be imposed by a competitive market for club memberships. When competition for membership exists, any club that did not provide maximum utility to its members, given the technology of providing the excludable public good, would not survive. Taking into account the budget constraint of a representative member, we obtain the following Lagrangian function to be maximized:

$$L = U(X, G, N) + \lambda(Y - P_x X - t). \tag{9.1}$$

If the club must operate under a balanced budget constraint, then t must satisfy $tN = F + P_g G$. Using this equation to replace t in (9.1), we obtain

$$L = U(X, G, N) + \lambda(Y - P_x X - F/N - P_g G/N). \tag{9.2}$$

Maximizing (9.2) with respect to X, G, and N yields first-order conditions

$$\frac{\partial L}{\partial X} = \frac{\partial U}{\partial X} - \lambda P_x = 0 \tag{9.3}$$

$$\frac{\partial L}{\partial G} = \frac{\partial U}{\partial G} - \lambda P_g/N = 0 \tag{9.4}$$

$$\frac{\partial L}{\partial N} = \frac{\partial U}{\partial N} + \frac{\lambda(F + P_g G)}{N^2} = 0. \tag{9.5}$$

From (9.3) and (9.4) we obtain

$$N\frac{\partial U/\partial G}{\partial U/\partial X} = \frac{P_g}{P_x}. \tag{9.6}$$

The quantity of public good provided to club members must be chosen so that the Samuelsonian condition for Pareto-optimal provision is satisfied; that is, the summation of the marginal rates of substitution of public for private goods over all club members must equal the ratio of their prices.

From (9.4) and (9.5) we obtain

$$N = -\frac{\partial U/\partial G}{\partial U/\partial N} \cdot \frac{F + P_g G}{P_g}. \tag{9.7}$$

If an expansion of club size induces unwanted crowding, $\partial U/\partial N < 0$, and (9.7) implies an $N > 0$. The larger the disutility from crowding relative to the marginal utility of the public good, the smaller the optimal club size. The greater the fixed costs of providing the public good to club members, the larger the optimal size of the club, owing to the advantages of spreading these fixed costs over a larger club membership.

The assumption that individuals have identical tastes and incomes is more than just an analytic convenience. It is often inefficient to have individuals of different tastes in the same club if this can be avoided. If all individuals are identical, except that some prefer rectangular pools and others oval ones, then the optimal constellation of clubs sorts individuals into oval and rectangular pool clubs.[2] Some differences in tastes for the public good can be accommodated efficiently in a single club, however. For example, if some individuals wish to swim every day and others only once a week, this heterogeneity of preferences can be efficiently handled by charging the different members different fees for the club service. If the only costs from increasing club size come from crowding, the optimal fees to finance the club will include a charge per visit. A similar user fee is needed to obtain the optimal allocation and use of the club good, if the costs of providing it (maintenance, for example) are positively related to use (Berglas, 1976; Sandler and Tschirhart, 1984, 1997, pp. 342–3; Cornes and Sandler, 1986, pp. 179–84).

If the constellation of preferences and technologies for providing excludable public goods is such that the number of optimally constituted clubs, which can be formed in a society of a given size, is large, then an efficient allocation of these excludable public goods through the voluntary association of individuals into clubs can be envisaged. Pauly (1967, p. 317) compares the rules or charter of the club to a social contract unanimously accepted by all members, and the theory of clubs, under these assumptions, is obviously much in the spirit of the contractarian and voluntary exchange approaches to public choice and public finance. With large numbers of alternative clubs available, each individual can guarantee himself the equal benefits for an equal share of the costs assumed earlier, since any effort to discriminate against him will induce his exit into a competing club, or the initiation of a new one. If optimal club sizes are large relative to the population, however, discrimination is possible and stable equilibria may not exist. With an optimal club size of two-thirds of the population, for example, only one such club can exist. If it forms, those not in it are motivated to lure members away by offering disproportionate shares of the benefits gained from expanding the smaller club. But the remaining members of the larger club are motivated to maintain club size, and can attract new members by offering the full benefits of membership in the big club; and so on. No stable distribution of club sizes and benefits need exist (Pauly, 1967, 1970). Analytically, the problem is identical to the emptiness of the core in the presence of externalities discussed in Chapter 2, or more generally the cycling problem (see Section 9.4*).

Even when a stable constellation of clubs exists, when optimal club sizes are large relative to the population's size, not all individuals may be part of an optimally constituted club. Although the voluntary association of individuals to form clubs increases their utilities, it may not maximize the aggregate utility of the entire population, defined to include those not a part of optimally sized clubs (Ng, 1974; Cornes and Sandler, 1986, pp. 179–84). We illustrate this point in Section 9.3 with a slightly different form of club.

9.2 Voting-with-the-feet

In the theory of clubs, exclusion from the consumption of the public good is assumed to be possible through some institutional device. A fence is built around the swimming pool and only club members are allowed inside the fence. Even when

[2] Buchanan (1965a) and McGuire (1974).

there is no fence around the swimming pool, however, those individuals who live a great distance from the pool are effectively excluded from its use by the costs of getting to it. When the consumption of a public good requires that one be at a certain location, distance can serve as an exclusionary device. If different bundles of public goods of this type are offered at different locations, a spatial division of the population into "clubs" of homogeneous tastes would arise from individuals choosing to reside in that local polity, which offered them their ideal constellation of public goods. No ballots would have to be cast. All preferences would be revealed through the silent voting-with-the-feet of individuals exiting and entering communities, a possibility first noted by Tiebout (1956).

In contrast to the disappointing promise of majority rule, the utopian quality of the unanimity rule, and the imposing complexity of the newer, more sophisticated procedures, Buchanan's clubs and Tiebout's voting-with-the-feet seem to accomplish the task of revealing individual preferences by the surprisingly simple device of allowing people to sort themselves out into groups of like tastes. The efficiency and mutual gain Wicksell sought from the unanimity rule in his voluntary exchange approach to collective action arise through the voluntary association of individuals in clubs or local polities.

Buchanan described the properties of a single club, and the optimality conditions [(9.6) and (9.7)] for membership in a single, isolated club. Tiebout described the process of voting-with-the-feet as one that could achieve Pareto optimality with respect to the entire population. But a local polity is a form of club, and clubs are a type of polity. Thus, conditions (9.6) and (9.7) must also hold for a single local polity, and a world of clubs must in principle offer the same potential as the Tiebout model does for achieving Pareto efficiency defined over the entire population. Moreover, any problems of stability or Pareto inefficiency that one can show exist with respect to one model, probably hold for the other.

The following conditions to ensure the global optimality of excludable public goods provision thus apply to both the clubs and voting-with-the-feet models:[3]

1. Full mobility of all citizens
2. Full knowledge of the characteristics of all communities (clubs)
3. Availability of a range of community (club) options spanning the full range of public good possibilities desired by citizens
4. Absence of scale economies in producing the public good and/or smallness of the optimum scale of production relative to the population size
5. Absence of spillovers across communities (clubs)
6. Absence of geographical constraints on individuals with respect to their earnings

Assumptions 1 and 6 are peculiar to the voting-with-the-feet model, but some sort of freedom-of-association assumption is certainly implicit in the clubs model if it is to produce global optimality. Some special difficulties with respect to assumption 6 are discussed below. Assumptions 1 and 5 tend to work at cross-purposes. The larger

[3] See Tiebout (1956), Buchanan and Wagner (1970), Buchanan and Goetz (1972), McGuire (1972), Oates (1972), and Pestieau (1977).

the community, the more costly it is to leave it, and the lower mobility is. Thus, exit is a more reasonable alternative from small than from large communities. On the other band, the smaller the community, the more likely it is that the benefits from the provision of any specific public good will spill over onto other communities and cause externalities across communities and non-Pareto allocations.

Assumptions 2 and 3 raise complementary issues. The basic argument assumes a full range of possible baskets of public goods available at the start. But how is this spectrum of opportunities established? Two possibilities come to mind: some central authority or auctioneer could set up different local communities and clubs with different baskets of public goods and inform all potential citizens of the characteristics of each community club. There are two obvious difficulties to this resolution of the problem, however. First, assuming a central authority knows what baskets of public goods must be supplied disposes of a large portion of the preference revelation problem, which the model is supposed to solve. If the central authority knew which people had which preferences, it could simply assign individuals to the appropriate club or local polity. Second, even if it is to some extent feasible, this solution to the preference-revelation problem violates the decentralized spirit of the Buchanan and Tiebout models.[4]

More appropriate is the assumption that entrepreneurs exist who create clubs and polities, where needed, for a share of the "profits" generated from providing a desired quantity or package of public goods. These clubs and polities could be set up on a not-for-profit basis, in which case the rewards to the entrepreneurial founders would, presumably, come in a nonpecuniary form, for example, the power and prestige associated with founding an organization. Tiebout uses the term "city managers" rather than mayors for the local polities' leadership, presumably in recognition of their entrepreneurial role. Frey and Eichenberger (1995, 1999) have recently emphasized the creation of public goods clubs, which they name functional, overlapping, and competing jurisdictions (FOJC) as a way of better matching public good supply and citizen preferences.

It must also be stressed that many goods with significant joint supply characteristics, but for which exclusion is practicable, are provided by profit-seeking entrepreneurs. Television program production and broadcasting are good examples of an activity with significant joint supply properties, but for which exclusion is possible with the help of scrambling devices and coaxial cables. Thus, one finds private firms offering packages of television programs for fees alongside publicly provided program packages. The former are basically consumption clubs formed to consume a particular bundle of television programs, while the publicly broadcast programs are available to citizens only near the points of transmission. Land developers receive an entrepreneurial return for the particular constellation of public and private good characteristics that they combine in the communities they create.

As always with a market-provided good or service, full Pareto optimality cannot be assumed unless the good is provided competitively. Moreover, the provision of excludable public goods by a monopoly raises efficiency issues that go beyond those

4 See Pauly (1970) and McGuire (1972).

that exist for a private good monopoly (Brennan and Walsh, 1981; Burns and Walsh, 1981). Nevertheless, the presence of many profit-making firms in competition with nonprofit clubs and local polities in providing excludable public goods (television, recreation and sports, education, travel, health care) attests to the importance of the entrepreneurial function in providing excludable public goods.

Although clubs can be single (swimming) or multiple (tennis, golf, and swimming) goods providers, local polities inevitably supply a number of goods and services and possess the potential for supplying many more. As the number of public good dimensions increases, the plausibility of assumption 3 declines. With one public good issue to decide, such as the proportion of tulips in the public square, 101 communities suffice to allow each individual to consume his optimal fraction of tulips to the nearest percentile. With two issues, the proportions of oaks and tulips, the number of communities needed to ensure Pareto optimality leaps to 101 squared. Each additional public good raises the number of polities required to a higher exponent. If the number of public goods is very large, one reaches a solution in which the number of communities equals the size of the population. Each community/individual becomes a polity with a basket of public–private goods (garden, woods) tailored to his own tastes, a possible consequence of the model that Tiebout himself recognized.[5]

9.3 Global optimality via voting-with-the-feet

Pareto optimality in a global sense requires that the incremental change in net benefits to the community that an individual joins equal the incremental loss to the community he leaves:

$$\sum_{i=1}^{n} \Delta U_A^i = -\sum_{i=1}^{m} \Delta U_B^i. \tag{9.8}$$

The change in utility of the nth individual to join community A is his total utility from being in $A(U_A^n)$, just as his loss from leaving B is his total utility in B, U_B^m. Equation (9.8) can thus be rewritten as

$$U_A^n + \sum_{i=1}^{n-1} \Delta U_A^i = U_B^m + \sum_{i=1}^{m-1} \Delta U_B^i . \tag{9.9}$$

In a world of pure competition, each factor owner's marginal product is the same in all industries and areas. If externalities and other market failures are not present, the welfare of others is unaffected by one's location. All ΔU^i are zero except for the moving individual, and he naturally locates in his most favored community. With public goods present, the ΔU^i for individuals in a community are positive for an additional entrant, as the total costs of the public good become spread over a larger number of individuals. A new entrant thus confers positive externalities for a community producing a pure public good. Alternatively, a new entrant can produce

[5] See also Pestieau (1977).

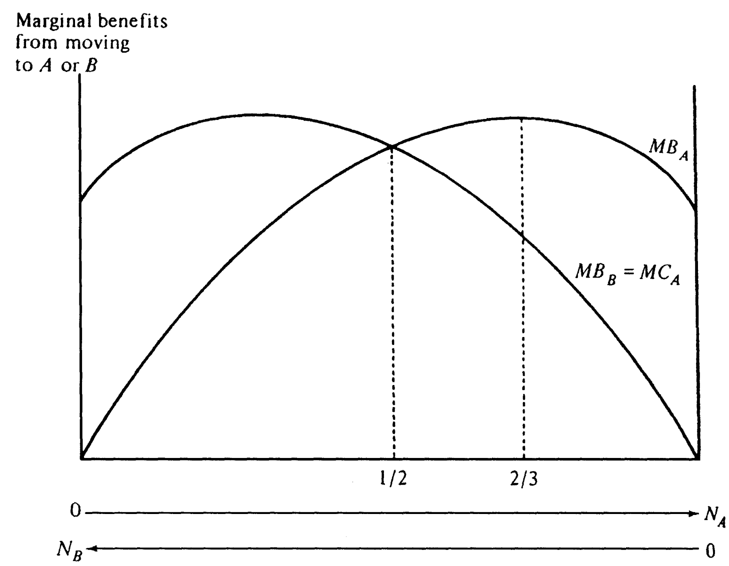

Figure 9.2. Marginal benefit from migration curves.

congestion costs, negative externalities, to a community that has grown beyond the optimal size for its locally provided public goods. In either case, since the moving individual compares only his utility levels in the two communities and ignores the marginal effects of his move on the others (the ΔU^is in A and B), voting-with-the-feet in general will not produce Pareto optimality in the presence of public goods and externalities.[6]

To see how a non-Pareto-efficient equilibrium can emerge, assume that there exist only two communities in which an individual can live, A and B. Each community is identical as are all of the residents. Each community provides a public good, which is optimally provided when two-thirds of the potential residents of the two communities consume it. Thus, there are enough individuals for only one optimally sized community. The situation is depicted in Figure 9.2. Curve MB_A represents the *average* benefits to a member of community A from membership in the community as a function of community size. These first rise as a result of the economies of scale property of the public good, and then begin to fall as crowding costs begin to outweigh the benefits from cost sharing. The curve MB_A also represents the *marginal* benefits to a member of community B from migrating to A. MB_B is the mirror image of MB_A defined with respect to the population of B.

The population of B is read from right to left along the horizontal axis. MB_B is also the marginal cost (MC_A) to a citizen of B from migrating to A. As usual, individual equilibrium occurs where marginal cost intersects marginal benefits *from below*. No such intersection exists in the figure. The intersection at an equal division of population is a local minimum. At any distribution in which one community has a higher population than the other, benefits are higher from membership in

[6] See Buchanan and Wagner (1970); Buchanan and Goetz (1972); Flatters, Henderson, and Mieszkowski (1974); and Pestieau (1977).

the larger community. Migration is from the smaller to the larger community, and this continues until all of the population enters into one of the communities. If congestion costs rise significantly, MB_A might decline fast enough following its peak to intersect MC_A. This would yield an equilibrium for the larger of the two cities at a size above its optimum, but below that of the entire population. In either case, however, the equilibrium city sizes achieved via voluntary migration are not those that maximize the average utility level of all individuals in the two communities. The latter would occur in this example when the population is equally divided between the two communities. This distribution of population maximizes the *average* benefits from *being* in either community. But, once this point is left, the *marginal* benefits from switching to the larger community exceed those of staying, and population redistributes itself until the stable but inefficient equilibrium is obtained (Buchanan and Wagner, 1970).

Although the assumption that the optimal-sized community is more than half the total population may seem unrealistic when one thinks of an area and population as large as the United States, often the potential migrant may not be considering such a wide spectrum of options. The relevant choice may be staying in small town B or moving to nearby large city A. Within this circumscribed range of choice, the optimal-sized community may be more than half the combined populations of the two communities, and the tendency for overpopulating the central city may become evident.

If the optimal-sized polity is less than half the population, the marginal benefits and cost schedules intersect and yield a stable equilibrium with the population evenly divided between the two communities. This equilibrium does result in a maximization of the potential benefits to each citizen, given the constraint that there should be but two communities. When addditional communities can be created, and the optimal-sized community is small relative to the total population, we return to a Tieboutian world in which free migration and the creation of new fiscal clubs can result in a set of communities, each of optimal size.

Additional complications are introduced into the Tieboutian world, however, if individuals earn part of their incomes outside the community. Assume again two communities with identical production possibilities, and individuals of identical tastes. Within any community, each individual receives the same wage, w, from supplying his services to the local production process, and a differential income, $r_i \geq 0$, that is tied to him and not to his location. This income can be thought of as coming from dividends, as in Tiebout's example, or as rents on assets peculiar to the individual, such as the income of a recording star. We shall refer to this income as simply rental income, covering all nonlocation-specific sources. Now consider two communities with equal numbers of workers, identical production possibility frontiers, and identical tax structures. In equilibrium the total private and public good production in community A must equal the sum of its rental and wage income:

$$\sum^{N_A} Y_i + G = N_A w_A + \sum^{N_A} r_i. \tag{9.10}$$

The utility of a resident of A is given by

$$U_i(Y_i, G, N_A), \tag{9.11}$$

as before. Substituting for G from (9.10) into (9.11), we have

$$U_i(Y_i, N_A w_A + \sum r_i - \sum Y_i, N_A). \tag{9.12}$$

The assumption of identical tax structures implies that the individual can purchase the same bundles of private goods, Y_i, in both communities. With equal populations and production possibilities, N_A and w_A equal N_B and w_B, respectively. Assuming the public good is not an inferior good, some of any additional rental income in A will go to increased public good production. Thus, $\sum r_i - \sum Y_i$ is larger in A than in B if $\sum r_i$ is larger in A than in B. Since public goods enter an individual's utility function with a positive sign, an individual is better off joining the community with the higher rents, assuming all other community characteristics are the same.

If the communities have different rental incomes, the same set of tax structures may not be optimal in both communities. Nevertheless, if tastes are the same, an individual always receives a more attractive tax–public good package from the community with higher rental income.

Higher rental incomes thus play the same role in attracting individuals from the other community as does a larger population in the presence of joint supply characteristics. Indeed, from (9.12) it can be seen that rental income, the wage rate, and the population size all enter the utility function in the same way through the public good term. Thus, any increase in population, the wage rate, or rental income, ceteris paribus, increases an individual's utility by increasing the quantity of public goods available. An increase in population also enters the utility function negatively, however, through the congestion effect represented by the third argument in the utility function. Increasing population can also be expected to drive the wage rate down, reducing an individual's command over private goods, and thereby his welfare. In contrast, higher rental income has an unambiguous positive impact.

Just as an individual's welfare is higher if he enters the community with the higher rental income, the community's welfare is higher, the higher the rental income of any new entrant. The depressing effects on wages and costs in terms of increased congestion from a new member are identical, but the benefits from increased tax revenue to finance public good provision are obviously greater, the greater the newcomer's rental income.[7] If the community has expanded to the point where the marginal gain from spreading the public good's costs over another taxpayer just equals the marginal cost in terms of reduced wages and congestion, adding another individual who is just a wage earner makes the community worse off. But if he has a high-enough rental income, the additional gains from financing an expansion of public good supply out of this rental income outweigh these costs. Regardless of what the community's size is, an additional member can always increase the welfare of all existing members if he brings with him a high-enough rental income.

[7] This effect is particularly apparent in the Flatters et al. (1974, pp. 101–2) model in which a golden rule is obtained where all rents go to public good production and all wages to private good production. This model is based on different assumptions from those of the discussion here, however.

In the same way that full mobility between communities may not bring about a Pareto-optimal distribution of the population, where economies of scale in public good production are large, full mobility is unlikely to bring about a Pareto-optimal distribution of the population in the presence of rents. In the preceeding example, the socially optimal distribution of the population is that which equates the marginal product of a worker in each community. This occurs at equal community sizes. But if the distribution of rents differs between the two communities, migration toward the community with the higher rents will occur. This migration will continue until the fall in marginal product and rise in congestion costs are large enough to offset the advantage that this community has from higher rents, and average utility levels in the two communities are equal.

To achieve the socially optimal distribution of population, taxes and subsidies must he levied on either residence in or movement in and out of a given community. One possibility is to grant a central authority the right to make transfers across communities. Such an authority would then determine what the socially optimum distribution of population was, and levy taxes and subsidies to achieve this optimum distribution. In the general case, the central authority would attempt to achieve the equilibrium condition given in (9.9). This requires a tax equal to $\sum_{i=1}^{n} \Delta U_A^i$ on community A if A is the community that is, or would become, too large, and a subsidy equal to $\sum_{i=1}^{m} \Delta U_B^i$ to community B if it would lose population. If the only difference between the two communities were the level of rental income, the policy would be simple to implement. The central authority would levy a tax on rental income in the community with higher initial rental income and offer a subsidy to the community with lower rental income to bring about equal rental incomes and populations in both communities.[8]

Alternatively, Pareto optimality can be achieved in a decentralized way, by granting each community the right to tax immigration and emigration. If the externalities for community A from immigration were positive, it could offer a subsidy to newcomers equal to $\sum \Delta U_A^i$ and levy an identical tax on emigration. If B did the same, all individuals would be forced to internalize the external costs their moving entailed, and Pareto efficiency would be obtained.[9]

Although these alternatives have identical efficiency outcomes, they differ both in spirit and in their equity properties. The latter weds Tiebout's decentralized voting-with-the-feet with the theory of exclusive clubs to produce a decentralized solution to the population allocation problem. The enactment of such a system of taxes and subsidies by local communities immediately provides communities favored by natural characteristics, population size, income, and so on with a valuable property right, which they exercise by taxing members outside their community (i.e., those who would have entered in the absence of the tax-subsidy scheme). The centralized solution vests the entire population with a property right in both communities and achieves allocational efficiency by taxing *all* members of the favored community to subsidize the disfavored community.

8 Flatters et al. (1974) and McMillan (1975).

9 Buchanan (1971) and Buchanan and Goetz (1972).

The difference in policies can be most easily seen by considering again our rent example but assuming that individual rents are not tied to given individuals, but are locational rents accruing to all residents of a given community. Granting the right to tax migration into the community with higher rents to its residents would allow them to achieve permanently higher utility levels than would be achieved in the less-favored community. Those who were lucky enough or quick enough to be born or move into a geographically more desirable area would forever be better off than those left in the less desirable areas. In contrast, the centralized solution would equate utility levels across communities by taxing the higher rent areas and subsidizing the low-rent ones.

Even when rental incomes are tied to individuals rather than locations, Tieboutian revelation of preferences coupled with local taxes and subsidies can raise equity issues. As noted earlier, a community can always be made better off by admitting someone whose rental income is high enough. Once a community has reached its optimal size for sharing the costs of public goods, it might adopt a policy of, say, admitting only new members who bring with them a rental income above the average. This can be accomplished by establishing zoning requirements on lot sizes and apartment dwellings that effectively screen out those with incomes below a given level. The mobile individual, on the other hand, is better off joining a community with greater rental income than he receives. The intersection of these two strategies could be a sorting out of individuals into communities of equal rental incomes. The identical incomes and preferences assumption that Buchanan assumed for convenience in initiating the study of clubs is a plausible outcome to a Tieboutian search for optimum communities (Buchanan and Goetz, 1972; Epple and Romer, 1991).

9.4* Clubs and the core

The preceeding discussion raises three issues with regard to the global properties of a world of clubs and voting-with-the-feet preference revelation: (1) whether an equilibrium distribution of the population among the clubs (communities) exists, (2) whether any equilibrium that occurs is Pareto efficient, and (3) what the redistributive-equity properties of the outcomes are. To further illustrate these issues we consider a simple example first presented by Ellickson (1973).

Assume that each individual i has the hyperbolic utility function $u_i = x_i g$ defined over private good x and public good g. Each individual in a club consumes the same quantity of g. Since $\partial u_i/\partial x_i = g$, the marginal utility of the private good is the same for all individuals within a club. We are working with transferable utility in x.

The unit costs of providing the good g to clubs of size 1, 2, and 3 are, respectively, a, b, and c. If $a = b = c$, we have a pure public good. If $a = 1/2b = 1/3c$ we have a pure private good. If the good is a pure public good, the optimal club size is the population. If it is a pure private good, optimal club size is one. We assume a public good with congestion costs so that

$$a < b < 2a$$

$$b < c < (3/2)b.$$

Consider first the quantity of g chosen and utility level obtained when an individual acts alone. Let w_i be i's wealth. We maximize u_i subject to the budget constraint $w_i = x_i + ag$; that is,

$$L_i = x_i g + \lambda(w_i - x_i - ag). \tag{9.13}$$

Maximizing with respect to g and x_i,

$$\partial L_i / \partial g = x_i - \lambda a = 0 \tag{9.14}$$

$$\partial L_i / \partial x_i = g - \lambda = 0. \tag{9.15}$$

Solving for x_i,

$$x_i = ag. \tag{9.16}$$

From the budget constraint and (9.16)

$$w_i = x_i + ag = 2ag, \tag{9.17}$$

from which

$$ag = \frac{w_i}{2} \tag{9.18}$$

and

$$u_i = x_i g = ag^2 = \frac{w_i^2}{4a}. \tag{9.19}$$

Equation (9.19) gives the security level of utility for any individual i, the level of utility i can achieve acting alone. No individual joins a club or community unless she can secure a utility of at least $w_i^2/4a$.

Let us now derive the conditions under which a club of two forms. The Samuelsonian condition for Pareto optimality requires that the sum of the marginal rates of substitution (MRS) for the two club members equals the marginal cost of the public good; that is,

$$MRS_i + MRS_j = b. \tag{9.20}$$

Now

$$MRS_i = \frac{\partial u_i / \partial g}{\partial u_i / \partial x} = \frac{x_i}{g}, \tag{9.21}$$

so that

$$\frac{x_i}{g} + \frac{x_j}{g} = b \tag{9.22}$$

or

$$x_i + x_j = bg. \tag{9.23}$$

The combined budget constraint for the club is

$$w_i + w_j = x_i + x_j + bg. \tag{9.24}$$

From (9.23) and (9.24) we obtain the Pareto-optimal quantity of the public good for a club of two.

$$g = \frac{w_i + w_j}{2b}. \tag{9.25}$$

To be induced to join a club of two, each individual must achieve a utility level of at least what she can achieve acting alone. From (9.24) we can write i's utility as

$$u_i = x_i g = (w_i + w_j - x_j - bg)g = (w_i + w_j)g - bg^2 - x_j g. \tag{9.26}$$

Now $x_j g$ is j's utility. If we set that at $w_j^2/4a$, the minimum level j is willing to accept and be in the club, then whether a club of two forms can be determined by seeing whether i's utility in the club exceeds her security level, that is, whether

$$u_i = (w_i + w_j)g - bg^2 - \frac{w_j^2}{4a} \geq \frac{w_i^2}{4a}. \tag{9.27}$$

Using (9.25) to replace g and some algebra yields

$$\frac{(w_i + w_j)^2}{b} \geq \frac{w_i^2 + w_j^2}{a} \tag{9.28}$$

as the necessary condition for a club of two to form. Whether a club forms depends on the respective wealth of i and j and the relative costs of supplying g in the two contexts. To see what is involved, assume that $w_j = \alpha w_i$, where $0 \leq \alpha \leq 1$. Then for (9.28) to hold the following condition must be satisfied:

$$\frac{1 + 2\alpha + \alpha^2}{1 + \alpha^2} \geq \frac{b}{a}. \tag{9.29}$$

Both sides of (9.29) lie in the range between 1 and 2, but the lower α is, the lower the left-hand side of (9.29) is. For a club of two to form, j's income must be sufficiently high relative to i's to allow her share of the costs of g to be large enough to compensate i for the crowding effect j's joining the club has (i.e., b's being greater than a).

The condition for the Pareto-optimal provision of the public good to a club of three requires that

$$g = \frac{w_i + w_j + w_k}{2c}. \tag{9.30}$$

In a manner analogous to the above demonstrations, one can show that the value of a coalition of three, $V(ijk)$, is $(w_i + w_j + w_k)^2/4c$. For the grand coalition to form, (9.31) and (9.32) must be satisfied:

$$\begin{aligned} V(ijk) &\geq V(i) + V(j) + V(k) \\ V(ijk) &\geq V(ij) + V(k) \end{aligned} \tag{9.31}$$

$$\begin{aligned} V(ijk) &\geq V(jk) + V(i) \\ V(ijk) &\geq V(ik) + V(j) \end{aligned} \tag{9.32}$$

where $V(i) = w_i 2/4a$, and $V(ij) = (w_i + w_j)^2/4b$. Suppose now that i and j have the same incomes, and k's income is α fraction of i's; that is,

$$w_i = w_j = w$$
$$w_k = \alpha w.$$

Consider just the implications of (9.32). Note first that an outcome in which i and j form a club dominates an outcome in which either i or j plays alone and the other forms a club with k:

$$V(ij) + V(k) \geq V(jk) + V(i) = V(ik) + V(j) \tag{9.33}$$

since

$$\frac{(2w)^2}{4b} + \frac{\alpha^2 w^2}{4a} > \frac{(1+\alpha)^2 w^2}{4b} + \frac{w^2}{4a} \tag{9.34}$$

if $b/a < 2$ and $\alpha < 1$. Thus, if only a club of two forms, it will be the wealthier two individuals that form the club. For the poorer k to be admitted, (9.35) must hold:

$$\frac{(2+\alpha)^2 w^2}{4c} > \frac{4w^2}{4b} + \frac{\alpha^2 w^2}{4a}. \tag{9.35}$$

The smaller c is relative to b and a, and the larger α is, the more likely (9.35) is to be satisfied. The poorer k will be invited to join the club by i and j if her income is high enough.

Now assume that $\alpha = 1/3, a = 1, b = 3/2$, and $c = 2$. Given these parameter values, (9.35) does not hold and a club of three does not form. A club of the two wealthier individuals will form, however, since $4w^2/4b > 2w^2/4a$ with $b = 3/2$ and $a = 1$. If the two wealthier individuals can both form a club and keep k out, they will. If, however, it is not possible to prevent individuals from moving into the community, k may choose to do so. Whether she chooses to join the community will depend on her assigned tax share once there. If, for example, the community were required to finance g by charging all members the Lindahl tax price for g, k would be better off in the community than if she remained outside and provided g for herself. Her Lindahl tax price is her *MRS*, which is x_k/g. Thus, from the budget constraint,

$$w_k = x_k + \frac{x_k}{g} \cdot g \tag{9.36}$$

or

$$x_k = w_k/2. \tag{9.37}$$

Half of k's income goes to pay for g, and half is left for private good consumption. Given her Lindahl tax share, her utility in the community of three is

$$u_k = x_k g = \frac{\alpha w}{2} \frac{(2+\alpha)w}{2c} = \frac{7}{72} w^2, \tag{9.38}$$

while playing the game alone she has only

$$u_k = \frac{\alpha^2 w^2}{4a} = \frac{w^2}{36}. \tag{9.39}$$

Thus, k will choose to join the community if she can, even though the aggregate utility of the community is lower with her in it than it is when she is outside it. It should also be obvious that k could choose to move to the richer community even if she left behind other k's who were made worse off by her departure from their community.

Even though the club of three provides lower aggregate utility than the club of two plus k playing the game alone, the effective redistribution from the richer two members to the poorer one when g is provided to all three members and financed at Lindahl tax prices makes her entry into the community to her advantage. We witness here exactly the same kind of Pareto-inefficient redistribution that we observed in Chapter 5 when a pure private good was provided to a community at equal quantities as *if* it were a public good and financed at Lindahl tax prices.

The Pareto inferiority of the club-of-three solution in this example implies that i and j would be better off bribing k to stay out of the community if her entry requires that she be charged only the Lindahl tax price for g. They even would be better off, of course, if they could prevent her from entering by forcing her to pay more than her Lindahl price, by charging her an entrance fee, or by some other institutional device (e.g., a zoning requirement).

Finally, we show that when the grand coalition is not in the core, no core may exist, even though a coalition of two can provide its members with higher utilities than when they play alone. Assume $w_i = w_j = w_k = w$. Let a, b, and c be such that

$$\begin{aligned} V(ijk) = \frac{(3w)^2}{4c} < \frac{4w^2}{4b} + \frac{w^2}{4a} = V(ij) + V(k) > \frac{3w^2}{4a} \\ \frac{3w^2}{4a} = V(i) + V(j) + V(k). \end{aligned} \tag{9.40}$$

At least one member of the $i - j$ coalition must pay at most her Lindahl tax price so that this individual's utility is at least

$$u_i = \frac{w}{2} \cdot \frac{2w}{2b} = \frac{2w^2}{4b}. \tag{9.41}$$

But (9.40) implies that

$$\frac{4w^2}{4b} > \frac{2w^2}{4a} = 2V(k). \tag{9.42}$$

Thus, a member of the $i - j$ coalition who pays at most her Lindahl tax price must have higher utility than the individual left outside of the coalition. The outside individual k must be able to offer the member of $i - j$ paying at least the Lindahl tax price a more attractive proposal to form a two-person coalition and $i - j$ cannot be sustained. We have here precisely the same kind of instability we confronted in Chapter 2 in the presence of multiple externalities (Aivazian and Callen, 1981).

9.5 Voting-with-the-feet: empirical evidence

In the Tiebout model rational individuals exit communities offering less attractive public goods–tax packages in favor of those providing more attractive packages. Three sets of testable implications follow from this assumption: (1) individuals do move in response to local government expenditure–tax offerings, (2) this migration process sorts people into groups of homogeneous tastes consuming the bundles of public goods of their choice, and therefore (3) individuals are more satisfied with their local public goods–tax packages where Tiebout sorting takes place.[10]

With respect to the first implication, numerous studies have found that both the levels of local public services and tax rates influence *whether* a family moves, and the choice of community into which it moves.[11] For example, an examination of responses to survey questions in the Columbus, Ohio, area in 1966 indicated a significant correlation between individual perceptions that there were problems facing the neighborhood and intentions to move (Orbell and Uno, 1972). Moreover, there was a greater tendency in urban areas to resort to exit instead of voice than there was in the suburbs. Individuals appeared to feel that voice is a more effective option in suburbs than it is in the city. John, Dowding, and Biggs (1995) report that a fifth of those who changed jurisdictions in the London area gave tax rates as an important factor in their decision to move.[12]

While the wealthy move away from high taxes, the poor move toward high welfare payments (Gramlich and Laren; 1984; Blank, 1988; Cebula and Koch, 1989; Cebula, 1991). So systematic is this migration, that state governments take it into account when setting welfare payment levels. A state that has large numbers of poor people living in its neighboring states sets a lower level of welfare payments (Smith, 1991).[13]

A particularly well-suited group for testing the Tiebout hypothesis is the elderly, since their incomes are typically from nonwage sources and thus their choice of residence is not likely to be dependent on characteristics of the job market. Cebula (1990) found that the elderly were significantly more likely to move to states that did not have an income tax. Results of Conway and Houtenville (1998), however, paint a much more complicated picture. They attempt to account for both the tax incentives that the elderly have to move from one state to another, and for the government expenditure incentives. Their results for out-migration generally support the predictions of the Tiebout model. Elderly citizens are more likely to leave states with high tax shares and high prices for public services. High property taxes appear

[10] A fourth possible implication, that housing values are bid up in high expenditure/tax communities (Oates, 1969) is more problematic and is not reviewed here. See, however, Edel and Sclar (1974); Hamilton (1976); and Epple, Zelenitz, and Visscher (1978).

[11] For a review of the literature up to 1979, see Cebula (1979). For updates, see Cebula and Kafoglis (1986) and Dowding, John, and Biggs (1994).

[12] See also the additional evidence presented by Dowding and John (1996).

[13] For a review of this literature with additional references see Brueckner (2000).

Table 9.1. *Frequency distribution of income homogeneity indexes, Los Angeles county municipalities, 1950, 1970*

	0.333–0.339	0.340–0.349	0.350–0.369	0.370–0.379	0.400+	Total
1950	25 (0.60)	5 (0.12)	5 (0.12)	3 (0.07)	4 (0.010)	42 (100)
1970 (old cities)	9 (0.21)	13 (0.31)	11 (0.26)	4 (0.10)	5 (0.12)	42 (100)
1970 (new cities)	1 (0.03)	9 (0.30)	12 (0.40)	1 (0.03)	7 (0.23)	30 (100)
1970 (all cities including 3 old 1950 cities for which 1950 data were missing)	12 (0.16)	22 (0.29)	23 (0.31)	5 (0.07)	13 (0.17)	75 (100)

Note: Percentages in parentheses.
Source: Miller, *Cities by Contract*, Cambridge, MA: MIT Press, 1981, p. 134.

to be a particularly strong stimulus to exit a state. Conway and Houtenville's results for in-migration do not support the Tiebout hypothesis, however. The elderly tend to move into states that have much the same characteristics as the ones that they leave. Factors other than the composition and efficiency of the public sector across states appear to determine the elderly's choice of a new home, once they choose to move.[14]

As in so many areas, California has led the world in the increasing trend toward greater mobility, with Los Angeles being the archetypical late-twentieth-century city. If the Tiebout process succeeds at sorting people into more homogeneous local communities, then the effects of the process should be apparent in Los Angeles. They are.

Gary Miller (1981, chs. 6 and 7) computed Herfindahl-like indices of income inequality (the sum of the squares of the percentages of the population in different income strata) for municipalities in Los Angeles County in 1950 and 1970. Since he used only three income strata, complete income heterogeneity would imply an index of 0.333, while complete homogeneity (all residents in the same income strata) would imply an index of 1.0. In 1950, 60 percent of the 42 cities for which data were available were virtually indistinguishable from the maximum degree of heterogeneity, and from Los Angeles County as a whole (index = 0.335) (see Table 9.1). Only 10 percent of 1950 municipalities fell into the most homogeneous category (0.400+).

The distribution of indices in 1970 shifted distinctly toward greater homogeneity, with only 16 percent of the municipalities in the most heterogeneous category and 17 percent in the most homogeneous category, even though Los Angeles County

[14] Part of the explanation for the inconsistency between Conway and Houtenville's results and the Tiebout hypothesis may be due to the level of aggregation of their analysis. States with high property taxes and levels of public services may on average attract individuals who make desirable neighbors. The local communities in which the elderly locate in these states may, however, have low education expenditures and property taxes. The elderly's choice of a new home may also simply be dominated by nonpublic-sector factors like the desire to be near children or grandchildren.

as a whole remained heterogeneous in income in 1970 (index = 0.334), as it was in 1950. Perhaps the strongest evidence that the Tiebout process does result in increased income homogeneity comes from the 30 newly created municipalities. To the extent that new municipalities come into existence to satisfy demands unmet by existing communities, their composition should accord most closely, in an age of high mobility, with the Tiebout hypothesis. Only 1 of the 30 newly created municipalities had income heterogeneity comparable to that of the entire county; almost one-fourth of the new municipalities fell into the most homogeneous category. In Miller's study, it appears to be largely a common preference for lower taxes and the avoidance of the redistributive outlays of the larger, older cities that drives the formation of new, suburban communities. Miller also presents evidence of increasing racial homogeneity within, and increasing heterogeneity across Los Angeles municipalities between 1950 and 1970.

Grubb (1982) also documents Tiebout sorting in the Boston metropolitan area, and Hamilton, Mills, and Puryear (1975) find less inequality of income within Standard Metropolitan Statistical Areas (SMSAs), the greater the number of school districts from which citizens can choose, and in general a better fit to Tiebout-model variables for suburban than for central city observations. Similar results have been reported by Eberts and Gronberg (1981). The Tiebout process is again found to work as predicted, and in so doing to produce less dispersion of incomes within the local polity.

Using a much longer time span than all the other studies, Rhode and Strumpf (2000), however, have found evidence of *decreasing* heterogeneity *across* communities using several measures of heterogeneity. Their work suggests that additional factors beyond mobility affect intra- and intercommunity heterogeneity over the very long run.[15]

Corroborative evidence of a different kind has been presented by Munley (1982) and Gramlich and Rubinfeld (1982a). Tiebout sorting should be more complete the greater the number of different political jurisdictions in which a mobile citizen can choose to live. Consistent with this prediction is Munley's finding that the dispersion of voter demands for education in Long Island, New York, decreased as the number of school jurisdictions in a geographic area increased. Similarly, Gramlich and Rubinfeld find a smaller residual variance in expenditure demands in the Detroit metropolitan area than in other parts of Michigan where a smaller number of local communities are available to the citizen.

Implicit in the Tiebout process is the assumption that when citizens with homogeneous preferences form a community, the community supplies the level of expenditures that these citizens demand, and thus that the citizens are more satisfied with the bundles of local public goods, which they consume. This part of the Tiebout model is supported by Gramlich and Rubinfeld's (1982a, p. 556) finding that two-thirds of Detroit metropolitan area voters surveyed wished to see no change in government expenditures, and the average desired change was only −1 percent. Although the percentage of voters desiring no change in expenditures (60 percent)

[15] Stein (1987) also offers equivocal evidence regarding Tiebout sorting.

was high throughout the rest of Michigan, that this percentage is lower than for Detroit suggests that the greater number of communities in which Detroit metropolitan area residents can choose to reside allows them to locate in communities that better provide them with the level of expenditures they demand.

Gramlich and Rubinfeld's findings are corroborated by Ostrom (1983) and Mouritzen (1989), who both report that citizens express greater satisfaction with local public services in urban areas with larger numbers of local jurisdictions. Brueckner's (1982) evidence that property values in 54 Massachusetts communities suggested neither an over- nor underprovision of local public goods provides still further support for the Tiebout hypothesis.

9.6 Voluntary association, allocational efficiency, and distributional equity

Wicksell's voluntary exchange approach achieves allocational efficiency by imposing a unanimity rule on the polity so that each collective decision must benefit all before it can pass. The approach assumes from the beginning that a predefined polity and citizenry exist.

The theory of clubs and voting-with-the-feet seek to determine a Pareto-optimal distribution of public goods through the voluntary association of individuals of like tastes. Here the dimensions of the polity and citizenship are outcomes of the "voting" process. These processes generally achieve Pareto optimality by grouping individuals into clubs and polities of homogeneous tastes. In the extreme, they satisfy Kramer's (1973) severe condition for consistent majority rule decisions, that all individuals have identical indifference maps, through the imposition of a silent unanimity rule.[16] These processes can realistically be assumed to come close to satisfying this goal, when, relative to the size of the population, (1) the number of public goods is small and/or (2) the number of distinct preferences for combinations of public goods is small. Since *the* task of public choice is the revelation of (differing) individual preferences for public goods, club formation and voting-with-the-feet, in part, solve the public choice problem by limiting its scope.

Despite these qualifications, the ability to exclude some individuals from the benefits of a public good remains a potentially powerful mechanism for revealing individual preferences. If A seeks the construction of tennis courts and B a golf course, then in a community in which all must consume the same bundle of public goods, and preferences are revealed by voting, regardless of what the eventual outcome is, it is likely to involve nonoptimal quantities of at least one good for one of the voters. This voter, say, A, is then worse off than she would have been had B also preferred tennis to golf and was willing to bear a larger share of this sport's costs. If A were incumbent to the community and B outside, A would clearly prefer that others with preferences closer to hers join the community, and, if it were in her power might discriminate in their favor over B.

[16] See also McGuire (1974), and on the relationship between voting-with-the-feet and the unanimity rule, see Pauly (1967, p. 317).

In Frey and Eichenberger's (1995, 1999) proposal it is not citizen mobility that drives competition across communities, but the entry and exit of political units in a federalist system.

None of this is very troubling if the public goods are tennis and golf, and the polities private clubs. No one objects too strenuously to a tennis club's restricting membership to those who want to play tennis. But the implications are less comforting for more general definitions of public goods. As we have seen, when individuals have positive income elasticities of demand for public goods, they can benefit from being in a community with incomes higher on average than their own from the additional units of the public good it provides. Even when each individual is taxed her marginal evaluation of the public good – that is, the Lindahl price – an effective redistribution from rich to poor occurs through the egalitarian distribution of the public good that of necessity occurs when rich and poor consume it together. But one's income elasticity of demand can be regarded as a sort of "taste" for a public good. If the incumbent membership of a local polity is free to exclude new members, then one can expect a sorting out of individuals into local polities of identical tastes *and* incomes, thus thwarting the possibility for this type of redistribution.

Wicksell assumed that voting on allocational issues took place following the determination of a just distribution of income. The same assumption could be made to support a voluntary association solution to the public good problem. But here it must be recognized that the voluntary association approach is likely to affect the distribution of income, while revealing preferences for public goods. A given distribution of private incomes might be considered just when individuals reside in communities of heterogeneous income strata, so that the relatively poor benefit from the higher demands for public goods by the relatively well-to-do. The same distribution of income might be considered unjust if individuals were distributed into communities of similar income and the relatively poor could consume only those quantities of public goods which they themselves could afford to provide.

The latter is the logical outcome of the voting-with-the-feet process, and one that is coming to pass. If the resulting distribution from this process were thought to be unjust, one could correct it by making transfers across communities, but here one runs directly into the issue of the proper bounds of the polity and the rights of citizenship.

In a federalist system there are two possible ways to view citizenship. Primary citizenship can reside with the local polity, and the central polity can be thought of as a mere union or confederation of the local polities with certain powers delegated to it. Conversely, primary citizenship can reside with the central state, with the local polities being merely administrative branches of the central government and having powers delegated from above. Under the first view of the polity, it would seem that the rights of the local polity to define its own citizenship and to pick and choose entrants would dominate the right of citizens in the larger confederation to migrate, free of hindrances, to any local polity. Here we see a direct clash between two of the conditions for achieving a decentralized, efficient allocation of public goods: the full mobility assumption, and the right of the local polity to tax and subsidize migration. If primary citizenship lies with the central state, then presumably individuals would be free to enter and exit local communities without incurring locally imposed penalties. Equity issues would be viewed from the perspective of the central polity, and it would be free to engage in intergovernmental transfers.

The same distinction exists with respect to clubs. The freedom to form voluntary associations can be regarded as one of the basic rights of the individual. To exercise this right in an optimal way, club members must be free to determine the quality and quantity characteristics of the excludable public good supplied to themselves *and* the size of the club's membership. When the supply functions for excludable public goods and the size of the population allow for the formation of many, individually optimally sized clubs, voluntary club formation can achieve a Pareto-optimal allocation of resources across the whole community. The outcome is entirely analogous to the Pareto-optimal allocation of resources that voluntary actions in the market achieve when large numbers of buyers and sellers exist. Indeed, firms are merely clubs of factor owners formed to achieve economies of joint supply in production, where the clubs discussed in this chapter arise to achieve economies from joint supply in consumption. Once again as in the market, however, when technology and population size combine to yield but a small number of optimally sized clubs, the independent utility-maximizing decisions of individuals may not achieve an outcome that is optimal from the perspective of the entire community.

In Chapter 2 we argued that the state emerges as a low-transaction-cost institution for achieving the cooperative agreements necessary for Pareto optimality in the presence of public goods and externalities. By extension, clubs, local polities, and the whole federalist institutional structure of the state might be formed to minimize the transaction costs of making collective decisions (Tullock, 1969; Breton and Scott, 1978).[17] But the discussion of this chapter reveals that the creation of new political jurisdictions within the state, the assignment of functions and revenue sources to different units, and the definition of citizen rights within a federated state raise issues that go beyond transaction costs savings and allocative efficiency. They go to the heart of the normative characteristics of the polity.

9.7 The theory of revolution

When neither the ballot nor the feet constitute adequate modes of expression, there is still Chairman Mao's barrel of the gun. One might expect to find more said about revolutions than has been the case, given their role in real-world politics. For the public choice analyst, the puzzle of revolutions is why individuals participate in them, and thus why they ever occur.

Consider the decision of individual i as to whether to participate in a revolution in her country, and if so how much time to contribute. She is unhappy with the present regime and anticipates benefits of β_i should the revolution succeed and a new order be imposed. The probability of this occurring is a function of the time i contributes to the revolution, t_{ir}, and the time all other citizens contribute, $O_{ir} = \sum_{j \neq i} t_{jr}$. Call this probability $\pi(t_{ir}, O_{ir})$. In addition to the gains, should the revolution succeed, i may receive personal pleasure from participating in the revolutionary movement, whether it succeeds or not, $P_i(t_{ir}, O_{ir})$.

[17] By further analogy, clubs of factor owners (firms) arise to minimize transaction costs in production (Coase, 1937).

Against these benefits the costs of participation must be weighed. Should i be caught and punished, she faces a fine or imprisonment promising a utility loss F_i. The probability that she will be caught, C_i, is a function of the time she devotes to the revolution, t_{ir}, the time others devote, O_i, and the resources expended by the regime to crush the revolution, R, that is, $C_i(t_{ir}, O_i, R)$ with expected partial derivatives

$$\frac{\partial C_i}{\partial t_{ir}} > 0, \qquad \frac{\partial C_i}{\partial O_i} < 0, \qquad \frac{\partial C_i}{\partial R} > 0.$$

In addition, by devoting time to the revolution, i foregoes income. If w is the market wage, then this opportunity cost is wt_{ir}.

The expected benefits from participating in the revolution are then

$$E_i = \beta_i \pi_i(t_{ir}, O_{ir}) + P_i(t_{ir}, O_{ir}) - F_i C_i(t_{ir}, O_i, R) - wt_{ir}. \qquad (9.43)$$

Maximizing (9.42) with respect to t_{ir}, we obtain

$$\beta_i \frac{\partial \pi_i}{\partial t_{ir}} + \frac{\partial P_i}{\partial t_{ir}} = F_i \frac{\partial C_i}{\partial t_{ir}} + w \qquad (9.44)$$

as the condition i must satisfy when determining her optimal level of revolutionary activity. The marginal expected gain in public good benefits (β_i) from an extra hour of participation plus the marginal personal enjoyment must equal the added risk of being caught when spending another hour in the revolution plus the foregone wage from not having worked that hour.

With O_i large, the change in both π_i and C_i from an additional hour of participation for the average person will be negligible. Whether someone participates or not, and if so to what degree, thus depends almost solely on the purely personal satisfaction from participation in the revolutionary movement weighed against the foregone income from taking time away from market activity (Tullock, 1971a, 1974), a result resembling that in the voting literature.

For the average citizen, the benefits from the revolution's success are the pure public good benefits from living under one regime rather than under another. But for a few, β_i represents the benefits from a position in the new government formed after the revolution. For these leaders, both β_i and $\partial \pi / \partial t_{ir}$ may be much larger than for the average individual. Thus, it is easier to explain the participation of the leaders of a revolutionary movement using a rational choice model than the participation of the rank and file (Silver, 1974; Tullock, 1974). Note, however, that for the leaders, F_i and $\partial C_i / \partial t_{ir}$ may also be higher. Under a rational choice theory, leaders of a revolution are like entrepreneurs in the theory of the firm, risktakers with extreme optimism regarding their ability to beat the odds.

The marginal effect of an average individual's contribution to the revolution's success should fall with the aggregate contributions of others, O_i. This free-rider effect will lower t_{ir} (Olson, 1965; Austen-Smith, 1981a). But there is also safety in numbers. The marginal risk of being caught, $\partial C_i / \partial t_{ir}$, may also shift downward with an increase in the revolutionary activity of others, thereby encouraging more revolutionary participation (Gunning, 1972; DeNardo, 1985). The personal rewards from participating in the revolution may also be characterized by a bandwagon effect and rise as others join the movement. Thus, participation levels could be characterized by increasing or diminishing returns to scale.

An increase in the resources devoted to crushing the revolution should lead to an increase in the marginal probability of getting caught, and thus discourage participation. Participation should be lower the higher the pecuniary costs, w.

Although the rational behavior approach to revolutionary activity gives some insights into why revolutions occur, it does not generate a rich harvest of testable implications. It does appear, however, that a revolution's success is greatly affected by the resources that the regime devotes to stopping it and thereby to curbing participation (Silver, 1974; DeNardo, 1985).

Perhaps the most distinctive implication of the theory is the prediction that participation declines with the wage rate. Austen-Smith (1981a) also has shown that it declines with a reduction in uncertainty about the wage if participants are risk averse. Tests of these implications by Finney (1987) indicate that the number of deaths from political violence in a country is negatively related to both the level and growth in national income, and is positively related to the standard deviation of the growth rate (a measure of uncertainty).

Although results such as Finney's are encouraging, it is yet to be seen how far a rational behavior model can go in explaining such extreme behavior as occurs in revolutions. Just as with the voluntary provision of a public good, the optimal choice for most rational individuals, when a meeting is called to stage a revolution, is to stay home (Olson, 1965; Dixit and Olson, 2000). Nevertheless, these models fill an analytical gap in the public choice literature. In a closed polity, an individual is always in danger of being "exploited" or "tyrannized" by a majority or minority of her fellow citizens. Her choices in such situations are to continue to rely on voice in the hope that the outcomes will change, to seek a new polity by migration, or to create a new one by revolution. The goal of public choice theory must be to explain all three choices.

Bibliographical notes

The discussion of efficiency and equity in a federalist system predates the public choice–Tiebout literature. See, for example, Buchanan (1950, 1952), Scott (1950, 1952a,b), and Musgrave (1961).

Ng (1985b) shows that one cannot have efficiency with club formation without violating *either* equity as argued earlier *or* freedom (voluntary association).

The clubs–Tiebout literatures have been surveyed by Henderson (1979); Sandler and Tschirhart (1980, 1997); Dowding, John, and Biggs (1994); and Inman and Rubinfeld (1997). Ostrom and Walker (1997) discuss the properties of a variety of club- and polity-like organizations.

The properties of markets in price-excludable public goods are analyzed by Oakland (1974), Burns and Walsh (1981), Brennan and Walsh (1981), and Walsh (1986).