1

The new economics of inequality and redistribution

Socialism, radical democracy, social democracy, and other egalitarian movements have flourished where they successfully crafted the demands of distributive justice into an economic strategy capable of addressing the problem of scarcity, and thereby promised to improve living standards on the average. Redistributing land to the tiller, social insurance, egalitarian wage policies, central planning, and providing adequate health care and schooling for all have been attractive when they promised to link a more just distribution of economic reward to enhanced performance of the economic system as a whole.

For this reason economic analysis has always been central to the construction of more democratic and egalitarian alternatives to capitalism, as well as to reforms of capitalism itself. Keynesian economics, for example, supported state regulation of the macro economy and also provided a rationale for income redistribution to the less well-off who, by spending a larger portion of their incomes, could be relied upon to generate higher and more reliable levels of demand for consumer goods, and thereby to sustain greater macroeconomic stability and higher levels of employment. Similarly, the model of general competitive exchange was deployed by socialists from Oskar Lange and Enrico Barone in the 1930s to Pranab Bardhan and John Roemer two generations later to demonstrate the possibility and advantages of democratic planning.

But today it appears that the left has run out of economic models. Keynesian policies to modulate fluctuations in aggregate demand are essential, but do not provide a foundation for a long-term egalitarian strategy. And while looming environmental catastrophe has underlined the need for public interventions to override the private-profit motive, centralized economic planning is incapable of regulating a complex, knowledge-based economy.

This is not to say that the left has abandoned the construction of alternatives to capitalism, as a reading of Bardhan and Roemer (1992), van Parijs (1995), Roemer (1996), Cohen (2009), and Wright (2010) will indicate. Nor have economists shrunk from the challenge of understanding the new global capitalist order (Glyn 2006, Bourguignon 2012) and designing policies to alleviate poverty (van Parijs and van der Veen 1986, Banerjee and Duflo 2011).

Yet even among egalitarians the conviction is widespread that while some combination of social democracy, market socialism, and workplace democracy would be preferable on democratic or egalitarian grounds to the capitalism we know, only capitalism has a workable answer to the problem of scarcity. Economic theory has proven, one hears, that any but cosmetic modifications of capitalism in the direction of equality and democratic control will exact a heavy toll of reduced economic performance.

Yet economic theory suggests no such thing. On the contrary, there are compelling economic arguments and ample empirical support for the proposition that there exist changes in the rules of the economic game which can foster both greater economic equality and improved economic performance. To see how this could be done, I need to explain how wealth inequality may be an impediment to productivity.

Inequality, institutions, and economic performance

First, some terms. *Co-ordination failures* occur when the independent actions of agents lead to outcomes less desirable

for some, and not better for anyone than could have been achieved in the presence of co-ordinated action. Economists term such an outcome *Pareto-inefficient*, meaning that there exists some technically feasible change in the current state such that some would be better off and none worse off. A *Pareto improvement* is a change that has this property; a *Pareto-optimal* state is one from which no Pareto improvements are possible. The latter is really a misnomer, because states with this benign designation may be highly unjust. (The terms are due to the Italian economist Vilfredo Pareto, 1848–1923.)

Examples of co-ordination failures are environmental pollution, unemployment, traffic jams, the creation of super-bugs through the misuse of antibiotics, spam, and the commonly observed inability of employers and workers to implement mutually beneficial changes in work rules and technology. The latter case – concerning employers and workers – is termed a *principal–agent problem*, the principal being the employer who pays a wage in return for the work time of the agent. Another important principal–agent interaction occurs in credit markets, where wealthy lenders (principals) lend money to borrowers (agents) in return for a promise of repayment with interest. Unlike traffic jams, in which all cars are (more or less) equal, principals and agents engage in asymmetric interactions: They differ in the actions each can take. The employer can offer a higher or lower wage, the worker can work hard or sleep on the job; the banker can charge a high or low interest rate, the borrower can repay or default.

These and other principal–agent relationships result in Pareto-inefficient outcomes. Compared to the situation in which both the employer and worker are taking the actions that maximize their objectives given the actions taken by the other (the *Nash equilibrium*), there exists a combination of a higher wage and greater work effort under which both the employer and the worker would be better off. And at the Nash equilibrium some would-be workers – even if identical

to those employed – will be without a job. Similarly (as we will see in the next chapter), some would-be borrowers will be excluded from the credit market entirely, even when the projects they would implement are superior to those being funded. Where credit transactions do occur, there exists a Pareto improvement over the Nash equilibrium: one in which the lender charges a lower rate of interest and the borrower takes fewer risks with his money.

Pareto-inefficient outcomes occur in other principal–agent relationships too, for example those between landlords and tenants (either agrarian or residential). Most relationships among people of different classes (in the traditional Marxian sense) are principal–agent relationships. Traffic jams and the threat of super-bugs are not.

But co-ordination failures indicated by these examples of Pareto-inefficient outcomes occur in principal–agent relationships for the same reason that traffic jams happen and superbugs proliferate. Co-ordination failures arise because some of the effects of an individual's actions on others – a more crowded highway, second-hand smoke, or a job well done, or the prudent use of borrowed funds so that repayment is assured – cannot be specified in an enforceable contract. The motorist who decides to drive downtown during rush hour cannot be charged for the additional congestion that she creates. The borrower's promise to repay will not mean much if he has gambled and lost it all. The source of the co-ordination failure in each case is not the absence of competition, or rigid wages, or "sticky prices," or "short-term maximizing," or any of the usual culprits. The problem is that the relevant contracts are incomplete. These and related cases are studied in detail in my microeconomics textbook (Bowles 2004). I use the broader term "co-ordination failures" (rather than the common "market failures") because, as these examples indicate, many of the failures take place in arenas other than markets.

The extent of co-ordination failures depends on what may be termed the *structure of economic governance*: the rules of

ownership, forms of competition, and norms and conventions that regulate the incentives and constraints faced by economic actors, and hence that determine the nature of co-ordination failures and their feasible solutions. The wealth of nations, as Adam Smith knew, depends critically on the structure of economic governance (or economic institutions for short); and the same can be said for the wealth of communities and firms (Acemoglu, Johnson, and Robinson 2005 and Acemoglu and Robinson, 2012). Ideally, a structure of governance is a means of avoiding or attenuating co-ordination failures, but there is nothing in the process determining the evolution of governance structures that insures this result. Governance structures may endure because they are favored by powerful groups for whom they secure a large slice of a given pie, not because these structures foster the growth of the pie itself.

The relationship between inequality and how productively a society uses its resources is thus mediated by the structure of economic governance. Governance structures also critically influence the degree of inequality. Correspondingly, the feasibility of distinct forms of governance is itself strongly influenced by the degree of inequality and, in particular, by the nature and distribution of property rights. For example, a co-operative-based governance structure in which those who supply labor to the production process also own the tools and equipment with which they work is hardly feasible where workers are very poor. A summary of the causal relationships between structures of governance, wealth inequality, and economic performance appears in Figure 1.1.

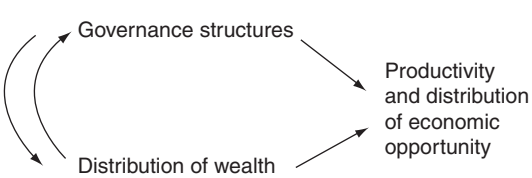

Figure 1.1 *Governance structure, wealth inequality, productivity, and inequality of opportunity*

I will define a change in governance structures as *productivity-enhancing* if the winners could compensate the losers (which would make the change a Pareto improvement), except that the implied compensation need not be carried out or even be implementable under the informational conditions and other incentive problems in the economy. The proposals developed in this book are motivated by the first key idea: *Inequality is an impediment to economic performance when it precludes implementation of productivity-enhancing governance structures*. There are three reasons why this is the case.

The first concerns the inefficient incentive structures that arise in economies with highly unequal asset distributions. An example may make this clear. Consider a single owner of a machine who hires a single worker to operate the machine who has no wealth. The worker has little reason to supply a high level of effort, since the worker is paid a given wage and the owner is the residual claimant on the income associated with the asset and hence receives the profit from the worker's labor. The *residual claimant* owns whatever remains (the residual) after all fixed claims (in this case the wage paid by the owner) are settled. Thus, without costly monitoring, productivity in the firm will suffer. But monitoring uses up resources that could have otherwise been productively employed. A rental contract in which the worker rents the machine from the owner for a fixed sum and becomes residual claimant on the entire income stream of the firm would of course avoid this particular incentive problem. But this solution to the effort–incentive difficulty simply displaces the conflict of interest to the issue of the treatment of the machine – in this case, the firm's capital stock itself. For the worker would then be residual claimant on the income produced by the machine, but not on the value of the machine itself, and hence would have little incentive to maintain the asset. Since the worker has no wealth, he or she cannot be the owner of the machine.

The generic problem here is that behaviors critical to high levels of productivity – hard work, maintenance of productive equipment, risk-taking, the production and use of knowledge and the like – are difficult to monitor and hence cannot be fully specified in any contract enforceable at low cost. As a result, key economic actors, workers and managers, for example, cannot capture the productivity effects of their actions as they would if, for instance, they were the residual claimants on the resulting income stream and asset value.

The result of these incentive problems is that a highly concentrated ownership of capital goods is often inefficient. We will see (Chapter 2) that there may exist a more egalitarian distribution, in which the worker becomes the owner of the firm's capital goods which, by more effectively addressing the incentive, monitoring, and maintenance problems involved, allows general improvements in well-being (including possible compensation for the former owner).

This being the case, one might wonder why the redistribution does not come about spontaneously. If worker ownership of the firm avoids incentive problems and supervision costs, it might be thought that owning the firm will be worth more to the worker than to the employer. But if this is the case, the worker would profit by borrowing to purchase the firm's capital stock. However, an asset-poor worker cannot borrow large sums (we will see why in the next chapter), and so he or she cannot purchase the firm's capital stock. Furthermore, the worker would be unlikely to agree to assume the risk of concentrated ownership of a risky asset, even if it could be financed. For this reason inefficient distributions of property rights – in this case the firm not being owned by the worker – may prove immune to disruption through private contracting despite the existence of other, more efficient distributions. More technically, inefficient property-right distributions may be sustained as a Nash equilibrium in a competitive equilibrium.

This one-worker firm example makes an important point, but it is unreal. Modern economies cannot avoid such incentive problems by implementing the simple property-ownership structures appropriate to an idealized Robinson Crusoe world of individual production. The economies of scale that characterize all contemporary economies make team production ubiquitous. In a capitalist firm the workers will shirk on the employer; in a co-op they will free ride on each other. These and related incentive problems will arise under any conceivable set of property distributions and institutional arrangements. So letting "the worker" own the machine is no magic bullet: co-ordination failures among a team of workers and (as we will see in the next chapter) their possibly over-prudent approach to risk-taking would have to be addressed. Nonetheless, differing levels of wealth inequality permit structures of economic governance that differ markedly in the costliness of the incentive problems to which they give rise, highly skewed wealth distributions supporting particularly inefficiency-prone governance structures.

A second reason why greater equality may enhance productivity arises because, where contracts are incomplete, the resulting co-ordination problems may be attenuated if people are intrinsically motivated to do a good job, to tell the truth, and to care about and to trust one another; and these sentiments are often difficult to sustain between the haves and the have-nots. Kenneth Arrow (1971:22) writes:

> It is useful for individuals to have some trust in each other's word. In the absence of trust it would be very costly to arrange for alternative sanctions and guarantees, and many opportunities for mutually beneficial cooperation would have to be forgone ... norms of social behavior, including ethical and moral codes [may be] ... reactions of society to compensate for market failures.

In addition to the invisible hand of competition and the fist of command, a well-governed society must also rely on the handshake of trust.

One of the possible productivity effects of greater equality may thus operate through the political and cultural consequences of redistribution. A well-run welfare state or a relatively equal distribution of property holdings may foster the social solidarity necessary to support co-operation and trust. These and related sentiments frequently provide the basis for low-cost solutions to co-ordination failures.

A third way in which equality may enhance productivity arises because institutional structures supporting high levels of inequality are often costly to maintain. Solving economic problems requires a state empowered to intervene effectively in the economy. But an activist state is capable of using its power not only to improve economic efficiency, but also to redistribute income in response to populist pressures. For this reason economic elites may prefer an ineffective state in an inefficient economy to a strong state in an efficient economy. Moreover, states in highly unequal societies are often obliged to commit a large fraction of the economy's productive potential simply to enforcing the rules of the game from which the inequalities flow: soldiers, police officers, prison wardens, and others in the ranks of what Arjun Jayadev and I call *guard labor* constituting large fractions of the labor force (Jayadev and Bowles 2005, Bowles and Jayadev 2007).

The private sector also incurs costs in enforcing inequality, in such forms as high levels of expenditure on work supervision and security personnel. Indeed, one might count high levels of unemployment itself as one of the enforcement costs of inequality, because the threat of job loss contributes to employers' labor discipline strategies. In less conflictual conditions, unemployed labor might be allocated to productive activities (we provide an illustration of how this might be done in Chapter 3). Moreover, in highly inegalitarian societies the insecurity of property rights is often widespread, militating against long-term investments by the rich and the poor alike.

Curious about the extent of and reasons for guard labor, Jayadev and I wondered if the demand for private guards was particularly high in US cities with very unequal distributions of income. Figure 1.2 shows what we found.

We also adopted a much broader concept of guard labor and sought to determine the amount of time devoted to the enforcement of claims on resources, including the protection of property rights and efforts to secure distributional advantage where contracts are absent or incomplete. We included supervisory labor, private guards, police, judicial and prison employees, military and civilian employees of the department of defense (and those producing military equipment), the unemployed, and prisoners. The data for the US are in Figure 1.3, and a cross-country comparison of the guard labor burden is in Figure 1.4. As in the case of private security guards in US cities, the extent of guard labor is correlated with measures of economic polarization (and also simply inequality of income), and varies inversely with measures of social welfare spending, as shown in Figures 1.5a–1.5b.

Where economic interactions are long on conflict and short on trust, technologies may also be chosen with the objective of improving an employer's bargaining power vis-à-vis his employees, reducing monitoring costs, or otherwise improving the labor discipline environment. Here is an example. When US trucking companies installed on-board computers during the 1980s, they vastly improved their ability to monitor the actions of the drivers (Baker and Hubbard 2000). Trip recorders provided the company with verifiable information on the speed, idle time, and other details of the operation of the truck about which there was a conflict of interest between the driver and the company. For example, the cost of operating the trucks (paid by the company) increased with the speed of the truck.

Drivers preferred to drive faster than the cost-minimizing speed, and to take longer breaks. Drivers who owned their trucks were residual claimants on their revenues minus these