The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models

Samuel Bowles

The American Economic Review, Volume 75, Issue 1 (Mar., 1985), 16-36.

Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198503%2975%3A1%3C16%3ATPPIAC%3E2.0.CO%3B2-N

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The American Economic Review is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aea.html.

The American Economic Review
©1985 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

http://www.jstor.org/
Wed Apr 23 14:54:42 2003

The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models

By SAMUEL BOWLES*

Recent years have witnessed a growing interest in the internal organization of the firm. Many, taking the work of Ronald Coase (1937) as their starting point, have developed insights based on the concept of transactions costs. Others, building on the work of J. R. Commons (1918, 1935), have developed an historical and institutional analysis of the structure of collective bargaining and internal labor markets. Others, starting from Marx's distinction between work ("labor") and labor time ("labor power") have developed an analysis of class conflict within the firm.

A careful reading of this diverse body of literature suggests that there are many common points of reference. All, for example, have stressed the social and nonmarket aspects of the production process.[1] But there are important differences as well.

In this essay I develop an underlying microeconomic logic of the Marxian model, and contrast it with two alternative views. The first is the simple Walrasian model in which the production process is represented as a set of input-output relations selected from an array of feasible technologies by a process of cost minimization with respect to market-determined prices. The Walrasian model presents no analysis of the internal social organization of the firm.

The second group of models stems from Coase's seminar work, and is exemplified by the important recent contributions of Armen Alchian and Harold Demsetz (1972), Oliver Williamson (1980), Guillermo Calvo (1979), Edward Lazear (1981), and others. Like the Marxian approach, and unlike the Walrasian, these models present a well-developed model of the firm as a social organization. I refer to these models as neo-Hobbesian because according to them the key to understanding the internal structure of the firm is the concept of malfeasance. Also known as shirking or free riding, malfeasance gives rise to the archetypal Hobbesian problem of reconciling self-interested behavior on the part of individuals with collective or group interests. Moreover, the neo-Hobbesian explanation of the functional nature of the hierarchical organization of the modern workplace bears a close resemblance to the original Hobbesian rationale for the state as a socially necessary form of coercion.[2]

By contrast, the basic commitment of the Marxian models is to the fundamental importance of class as an economic concept. While the Marxian model does not deny the importance of the Hobbesian conflict between individual and collective rationality as an underlying social problem central to an understanding of the production process in any social system, it focuses on those problems which may be traced to the structure of

*Department of Economics, University of Massachusetts, Amherst, MA 01003. I have benefited greatly from the comments and criticisms of my colleagues at the University of Massachusetts and the University of Siena, particularly Robert Costrell, Kenneth Flamm, Herbert Gintis, Richard Goodwin, Donald Katzner, Michael Kruger, Fabio Petri, Ugo Pagano, and Leonard Rapping. I also thank David Gordon, Robert Boyer, Frank Hahn, James Malcomson, Robert Solow, Juliet Schor, Robert Gordon, Duncan Foley, and two anonymous referees for criticism and comments, and the John Simon Guggenheim Foundation and the German Marshall Fund of the United States for financial support.

[1] The list of approaches is quite partial as it excludes, for example, the interesting and related work on social norms and economic processes. See George Akerlof (1980) and Robert Solow (1980).

[2] Thus, for example, it is argued that a team of workers would rationally hire a supervisor to monitor their work activities, an economic analogue to the Hobbesian position which asserts that uncoerced citizens in a state of nature would in their own interests commit themselves to obey the dictates of a state.

ownership and control of the means of production.[3]

What is at issue between Marxian and non-Marxian economists is not the general relevance of class concepts to the analysis of social groupings, institutions, or political action, but the status of class as an *economic* concept. Even within the realm of economics, terminological differences aside, there is general agreement on the relevance to a wide range of issues of what Marxian economists would term the class structure. Few economists of any persuasion would question the importance of the distribution of the ownership of assets as a determinant of the distribution of income, patterns of consumption, or levels of saving.

The Marxian model is distinct, however, in that it asserts that consideration of the ownership of the means of production, and the command over the production process which this ownership permits, is essential to a coherent analysis of the production process itself, and to the analysis of market equilibration and competition. It is thus not only in its macroeconomic theory and its theory of collective action that the Marxian model makes substantive use of the idea of class, but in its microeconomics as well.[4]

The distinctiveness of the Marxian microeconomics with respect to the neo-Hobbesian and Walrasian approaches, as we shall see, has little to do with the labor theory of value, however. Its primary focus is on the interactions between the voluntary relations of the marketplace and the command relationships of the workplace. Thus Marxian economists take strenuous exception to Paul Samuelson's assertion that "in the competitive model it makes no difference whether capital hires labor or the other way around" (1957, p. 894).

[3] I will specify what I take to be the principal differences between the neo-Hobbesian and the Marxian models in the penultimate section. The relationship between the Marxian model and what Marx wrote is suggested in various notes.

[4] I would thus take strong exception to Oskar Lange's (1935) view that the specificity and strength of Marxian economics resides in its institutional and sociological content and not in its microeconomic theory per se.

The structure of the Marxian model may be illustrated by reference to three propositions central to its analysis of capitalist production.

First, capitalists (owners of firms or their representatives) will generally select methods of production which forego improvements in productive efficiency in favor of maintaining their power over workers. For this reason, the technologies in use in a capitalist economy, as well as the direction of technical change, cannot be said to be an efficient solution to the problem of scarcity, but rather, at least in part, an expression of class interest. This proposition is fundamental to the Marxian assertion that the productive potential of a society (the "forces of production") is inhibited (or "fettered") by the specifically capitalist institutional structure of the economy (the "social relations of production").

Second, it will generally be in the interest of capitalists to structure pay scales and the organization of the production process to foster divisions among workers, even to the extent of treating differently workers who are identical from the standpoint of their productive capacities. This proposition is central to the Marxian divide and rule interpretation of internal labor markets, segmented labor markets, and discrimination.

Third, involuntary unemployment is a permanent feature of capitalism central to the perpetuation of its institutional structure and growth process. In a capitalist economy, product and labor markets will not function so as to eradicate Marx's familiar "reserve army of the unemployed." Moreover, even public policy towards this objective will be unable to maintain full employment.

To economists trained in the Walrasian or more generally neoclassical tradition, these assertions are often thought to be either nonsensical, or based on a radically different model of production and competition. Specifically, it is often thought that these propositions require one or more of the following assumptions: that capitalists collude in pursuit of their collective interests, that capitalists do not maximize profits, that product and factor markets are not competitive, or that the economy is characterized by im-

portant institutional rigidities such as sticky wages. Under these assumptions, it is not difficult to demonstrate the above propositions and thus affirm the importance of the Marxian concept of class.

But, while sufficient, these assumptions are not necessary to the demonstration of the above basic propositions of Marxian economics. (Nor, one might add in passing, are they particularly central to Marx's own theoretical writings, which generally presumed a highly competitive economy based on profit maximization.) The basic difference between the Marxian and Walrasian models is thus not in the structure of markets or in concepts of collective vs. atomistic action, or in institutional rigidities, but in the analysis of the process of production itself, or in what Marxists term the labor process.[5]

In this essay I develop a simple model of the production process in a competitive capitalist economy. To the familiar two-equation Walrasian model of production (production function and cost function), I add a third equation representing class conflict within the production process. I then derive the above three propositions from the expanded model. I close with some observations on the closely related but quite distinct neo-Hobbesian model of the production process.

My intent is not so much to advance the discussion of technical change, discrimination, or involuntary unemployment per se, as to provide a single coherent microeconomic framework capable of integrating important modern Marxian contributions in these fields. To cite only a few: those of Stephen Marglin (1974), William Lazonick (1982), and Harry Braverman (1974), on technology; of Richard Edwards, David Gordon, and Michael Reich (1982), Herbert Gintis (1976), and John Roemer (1979), on divide and rule strategies; and of Michel Kalecki (1943), Andrew Glyn and Robe Sutcliffe (1972), Raford Boddy and Jam Crotty (1975), and Richard Goodwin (1967 on unemployment.

[5] Partly as a result of the differing treatment of the labor process and partly for other reasons, the Marxian and Walrasian views of the competitive process differ somewhat. Both stress the importance of unlimited entry and a multiplicity of buyers and sellers. Marxists, however, generally assume price-making rather than price-taking behavior by firms.

I. The Extraction of Labor from Labor Power

The Marxian model comprises an analys of three quite distinct aspects of the produ tion process, broadly construed: market e changes (modeled as voluntary contractu or contract-like interactions), physical inpu output relations (which in principal might b represented by an engineering productio function), and social relationships amon workers and between workers and their em ployer (which are modeled in an entirel different manner).

Central to the Marxian approach is th distinction between those social relationship that take the form of market exchanges be tween firms and other ownership units, o the one hand, and relationships of comman that take place within firms. The marke arena in which contractual exchanges tak place, Marx termed "a very Eden of th innate rights of man." By contrast the inter nal structure of the firm—which Mar termed the "hidden abode of production"— is represented (as Coase was later to do) as mini-command economy.[6]

The distinction between the two types o social relationships would be of little theoret ical importance, of course, if the comman relations of the firm were simply effects en tirely derived from the technological struc

[6] The distinction is perhaps the most fundamental in Marxian economics. Marx wrote:

> If we consider the exchange between capital and labor, then, we find that it splits into two processes which are not only formally but also qualitatively different...: (1) the worker sells his commodity...(labor power)...which has...as a commodity...a price.... (2) The capitalist obtains labor itself...he obtains the productive force which maintains and multiplies capital.... The separation of these two processes is so obvious that they can take place at different times and need by no means coincide. The first can be and usually, to a certain extent, is completed before the second even begins.... *In the exchange between capital and labor the first act is an exchange and falls entirely within ordinary circulation; the second is a process qualitatively different from exchange, and only by misuse could it have been called any kind of exchange at all.* [1973, pp. 274–75]

ture of production and the market relationships into which the firm enters. Indeed, this is precisely the logic of Samuelson's remark quoted above.

But, according to the Marxian model, the structure and effects of the social relations within the firm—of command, cooperation, competition, and the like—while influenced by technology and market relations, are not entirely reducible to them, but rather depend on the class structure of the productive process, and hence require a distinct form of modeling. By contrast, Walrasian theory denies the need for a distinct modeling of the social relationships within the firm, while the neo-Hobbesian approach insists that a distinct modeling of the firm as a command economy is necessary, but has nothing to do with the class structure, for hierarchical relationships between managers and workers reflect nothing more than an efficient solution to the universal problem of malfeasance.

The importance of the social structure of the firm, the necessity of a distinct modeling of these social interactions, and the centrality of the class structure to their analysis may be traced within the Marxian model to three characteristics of the production process. First, labor is embodied in people, and hence labor services are inseparable from the person supplying the service. Second, whether for reasons of technology or of economies of supervision, production is generally less costly when it is done by a considerable number of workers together in one location. And third, the production process is always a process of joint production, as the workers' attitudes, capacities, and beliefs are transformed in the production process as surely as the raw materials and other goods in process are transformed into final outputs. I will refer to these three characteristics respectively as the human embodiment of labor, the social nature of production, and the endogeneity (or joint production) of workers.

Two types of social interaction within the firm are central to understanding the production process: relations among workers (of competition, solidarity, or whatever) and relations between workers and their employer. I focus on the second at the outset, representing the capital-labor relationship as a simple bilateral relationship between two individuals. Relations among workers will be introduced later.

The relationship between workers and their capitalist employer is formally structured by the ownership and control of the means of production. It is thus (by definition) a class relationship. In what follows, two characteristics of this relationship will be central. Both may be considered axioms with respect to the proposition to be demonstrated below. First, quite apart from the level of wages, employers and workers have a conflict of interest in the production process in the specific sense that the employer's interests (as measured by profits) are enhanced by being able to compel the worker to act in a manner that he or she otherwise would not choose. This conception of a conflict of interest does not imply that the employer and the worker have no common interests, or that, if left to their own devices, labor would choose not to produce anything at all. It simply states that within a given legal and economic context, the employer can do better than to simply hire workers and let them work as they please. The level of profits therefore depends—at least to some extent—on the power of capital over labor.

While this conflict of interest may extend to such issues as the safety or comfort of the workplace and the amount, type, and location of new investment, I focus in what follows on the conflict over the amount of work done per hour, or what may be termed the intensity of labor. This is often termed the conflict over extraction of labor from labor power. It might better be called the extraction of work from the worker.

The second axiomatic characteristic of the capital-labor relationship is that the strategies that capital may adopt in order to enhance or exercise its power over labor are costly. The basis of the power of capital over labor is the ability of the owner to impose costs on workers who refuse to (or otherwise fail to) carry out the wishes of the employer. In liberal capitalist societies, the only means by which this cost may be imposed is via the employer's control over the terms of employment (wage and other conditions) and the possibility of job termination. For reasons of

simplicity, I focus initially on the threat of job loss.

The expected cost to the worker of resisting (or otherwise not carrying out) the command (explicit or implicit) of the employer will depend on the likelihood that the worker's resistance will be detected, and on the cost to the worker of losing his or her job. (Assume for the present that any worker who is observed performing below the employer's expectation will be fired; I will later modify this assumption.) Because the cost (to the worker) of job loss will depend on the wage, enhancing the threat of job loss (by raising the wage) will be costly to the employer. Similarly, the employer cannot costlessly know what each worker is doing at any given moment even if the employer knows all of the workers' production capacities and personality characteristics. However, the employer can increase the probability of detecting below-standard work intensity through employing surveillance personnel and equipment, and by using production methods that produce (as a joint product) information on individual worker performance. Both methods of enhancing the worker's expected cost of working below expectation are thus costly to the employer.

These two characteristics of the production process—the conflict of interest between capital and labor, and the costliness of employer strategies—form the basis of the propositions that follow. The underlying reasoning may be made more precise with the aid of a simple model.

Let us assume that labor is homogeneous, that the employed and unemployed are otherwise indistinguishable, that there are no employer costs of selection or on-the-job training, that workers are risk neutral, and that all markets are competitive in the sense of a multiplicity of noncolluding buyers and sellers.[7]

Let the output of a firm be a function of the level of inputs.

$$Q = f(X, L), \tag{1}$$

where Q is the number of units of output over some period of time, X is the vector of material inputs and services, and L is the input of labor over this same time period. All inputs and output are measured in physical terms. Labor is thus counted in effective work done, or effort units. For simplicity, the price of the output is taken by the firm as given and is set equal to one.

As is quite evident, the treatment of total sales and the physical input-output aspect of production in the model is similar to its neoclassical—or Walrasian—analogue. The difference emerges when we consider the cost function. The labor argument in the production function—work effort—bears no market price, for it is labor time, not work itself, that is purchased. Hence the cost of labor—work—cannot be expressed in the firm's cost function as a market-determined hourly wage rate multiplied by the number of labor hours hired.[8] To express the cost function and the

[7] Unlike search models or Arthur Okun's (1981) toll model, I assume that workers have complete information about job and wage conditions throughout the economy, that employees know all (actual and potential) employee characteristics, and that what Okun called "the attachment between employer and employees (mutual),... the key component of the toll model that was absent in the simple search model" (p. 75) is absent here as well. Unlike contract theory, I assume away problem of risk aversion and issues of reputation (workers and capitalists alike have no memories).

[8] Marx (1976) dramatized the fact that labor itself cannot be bought and hence has no price as follows. "On the surface of bourgeois society the worker's wage appears as the price of labor, as a certain quantity of money that is paid for a certain quantity of labor" (p. 675). But "it is not labor which daily confronts the possessor of money (the capitalist, SB) on the commodity market, but rather the worker. What the worker is selling is his labor power" (p. 677). As a result, "according to the amount of actual labor supplied every day, the same... wage may represent very different prices of labor, i.e., very different sums of money paid for the same quantity of labor" (p. 683). Marx then makes it clear that the cost of a given amount of labor may vary through the extension of the length of the working day, or through an increase in the intensity of work in any given hour. "The rise in... wages may therefore be unaccompanied by any change in the price of labor, or may even be accompanied by a fall in the latter" (p. 684). Henry Ford may have understood this when he paid his workers in Detroit the unheard of sum of $5 a day. That labor itself cannot be purchased has long been recognized outside the Marxian tradition as well. Gary Becker observed that "any enforceable contract could at best specify the hours required on a job, not the quality of the performance" (1962, p. 6). But this fact has not been given the importance it has received among Marxian economists.

production and total sales function in the same terms, a third equation is required—the labor extraction function—representing the amount of labor done per hour of labor hired as a function of the costly inputs used to elicit work from workers.[9]

We may write L, the total labor input, as the product of the hours of labor power hired, Lp, and the amount of work done per hour l^*, or $L = Lpl^*$. The amount of work done per hour is determined by the worker in response to the constraints devised by the employer, given the availability of other jobs, unemployment insurance, and the like. At this point, attention need only be given to those determinants of the worker's effort that appear as instruments from the standpoint of the employer.

The amount of work done per hour will depend upon the worker's perception of the cost of pursuing a nonwork activity, that is, of acting on the basis of any of his or her nonwork (and work-reducing) objectives. Assuming that a worker's job will be terminated if the worker's nonwork activities are detected, the expected cost of pursuing nonwork activities, $E(n)$, is the product of two terms: the probability that a worker's nonwork strategy will be observed by the employer, p^o, and the cost of being fired, if observed, w^*. It is assumed that p^o is positively affected by the amount of surveillance inputs (material or human) purchased per hour of production labor hired, s, or $p^o = p^o(s)$, and $p^o(0) = 0$, and $p_{os} > 0$ for $s > 0$. (Here and below subscripted functions indicate the partial derivative of the function with respect to the variable indicated by the subscript.)

Surveillance labor does not enter into the transformation of inputs into outputs, and is thus distinct from what may be termed coordination labor, which is a production input represented in the production function as a component of L. (Here I abstract from the far from trivial problem of extracting work from surveillance employees. Thus I represent surveillance services, s, as purchasable at price p_s.) The cost of an hour of labor power, c_{Lp}, is thus $(w + p_s s)$, and the cost of an effort unit of labor, c_l, or what Marx called the price of labor, is $(w + p_s s)/l^*$.

The money cost of being fired is measured by w^*, the difference between the wage offered and the worker's expected income if fired. (I assume for simplicity that the worker has no nonwage income if employed.) This latter term is simply a weighted average of w^c, the worker's nonwage income if fired and not reemployed (unemployment insurance, means-tested income support payments, and the like), and $\hat{w}$ the expected wage in some other job, should the fired worker find employment elsewhere. It is assumed that both wages $(w, \hat{w})$ exceed w^c. Thus assuming a time horizon of a single period and letting j represent the probability of finding another job (or equivalently, the fraction of the period during which the worker expects to remain unemployed), the expected income loss, $\hat{w}^d$, is

$$\hat{w}^d = w - \left[j\hat{w} + (1 - j)w^c \right].$$

All of these wage terms, including w^c, are expressed in real units.[10]

A particularly simple model of the worker's response to the employer's choice of various

[9] Note that if labor costs did *not* depend on hours of labor hired but only on the amount of labor done, or if the relationship between hours hired and work effort performed were exogenously determined, or if the extraction of work from workers were costless, the third equation would be unnecessary. However, even the use of straight piece-rate payments will not render costs independent of the hours of labor hired unless the piece-rate workers use no inputs owned by the firm, and the determination of the number of pieces produced requires no surveillance inputs and hence is costless. But in this extreme case, there is no reason—by conventional definitions—to consider the piece-rate workers part of the firm that purchases their output, for their sole relationship to the firm is an exchange. The necessity for the third equation is thus based on assumptions no different from those used in the Coasian tradition to explain the existence of firms. The manner in which this function is developed is quite different, as we shall see, from its Coasian analogue.

[10] Note that because the employer clearly may directly set only nominal variables, but seeks to implement a real strategy, the general price level will enter into the employer's wage setting even in the absence of cost-of-living provisions in contracts. But I will not develop this point here.

combinations of surveillance and wage-loss threat results if we assume that at any moment the worker's decision is to work at a level of intensity satisfactory to the employer or not to work. The intensity of labor, l^*, then is just the percentage of time on the job during which the worker is actually working. It is assumed that the worker chooses a desired level of l^*, and then selects the moments of work and nonwork randomly. The probability that the worker will be detected not working, and hence dismissed, (p^d), is equal to the probability of being observed at any moment (p^o), multiplied by the probability that at that moment the worker will not be working ($1-l^*$), or $p^d = p^o(1-l^*)$. The probability of job retention is simply $(1-p^d)$, setting aside reasons for job termination other than observed nonwork. Thus, for $l^*=1$, $p^d=0$.

Let us assume for simplicity a two-period framework in which hiring occurs only at the beginning of a period and firing occurs only at the end of a period. The worker's time preference is assumed to be zero. The worker's expected income over two periods is thus the first-period's (assured) wage plus the expected wage or nonwage income for the second period:

$$\hat{y} = w + (1-p^d)w + p^d(j\hat{w} + (1-j)w^c).$$

Assuming identical workers and employers makes it reasonable to represent the worker as perceiving the alternative wage as identical to the present wage, or $w=\hat{w}$, and thus the expected income in the second period, if dismissed at the end of the first period would be $w-\hat{w}^d$, and rewriting the above expression for $\hat{y}$:

$$\hat{y} = 2w - p^d\hat{w}^d. \tag{2}$$

The worker's expected effort over two periods is both the effort expended in the current job, and the effort expended in the next job, should the worker be terminated and then reemployed. (Given the assumption that the worker has full information and hence nothing to learn, it is reasonable to suppose that the worker's choice concerning work effort when reemployed will be identical to the prejob loss choice.) Thus, the expected level of effort is

$$\hat{l}^* = l^* + (1-p^d)l^* + p_d j l^*. \tag{3}$$

The worker values income and, on the margin at least, finds increased work intensity displeasing.[11] The risk-neutral worker's response to the employer's strategy will be that which maximizes

$$\hat{u} = \hat{u}(\hat{y}, \hat{l}^*) \tag{4}$$

by equating the expected marginal disutility of effort (from equations (3) and (4)) with the expected marginal utility of income associated with an increment of effort (from equations (2) and (4)).[12]

Because the expected marginal income return to an increment in work will depend positively on $\hat{w}^d$, under quite general assumptions it can be shown that the worker's choice of l^* will be a positive function of $\hat{w}^d$.[13] By similar reasoning it can be shown that work intensity will be a positive function of s.

We may now represent the amount of work done per hour of labor power pur-

[11] This does not require a marginal disutility of labor (or effort). Even on the margin, the worker may enjoy the process of work, or despise it; what is essential to my argument is the assumption that the workers' objective function includes some positively valued on-the-job activities (or inactivity) that are associated with a positive opportunity cost in terms of working.

[12] That is, by equating $(\partial u/\partial \hat{y})(\partial \hat{y}/\partial l^*)$ with $-(\partial \hat{u}/\partial \hat{l}^*)(\partial \hat{l}^*/\partial l^*)$.

[13] Assuming the second-order conditions for the worker's utility maximization to be met, it can be shown that effort will be an increasing function of $\hat{w}^d$ for $\hat{w}^d \geq 0$, $l^* < 1$, and $s > 0$. This is because an increase in $\hat{w}^d$ will increase $(\partial \hat{u}/\partial \hat{y})(\partial \hat{y}/\partial l^*)$. This follows readily from the independence of $\partial \hat{u}/\partial \hat{y}$ from $\hat{w}^d_*$ and the fact that $\partial \hat{y}/\partial l^* = p^d \hat{w}^d$. Thus, $(\partial^2 \hat{y}/\partial l^* \partial \hat{w}^d)$ must also be positive (for $s, \hat{w}^d > 0$). The upward shift of $(\partial \hat{u}/\partial \hat{y})(\partial \hat{y}/\partial l^*)$ associated with an increment in effort will necessarily result in an increase in effort as long as the disutility associated with a marginal increment in effort is not infinite. Thus the derivative of work effort with respect to the cost of job loss will be positive for positive $\hat{w}^d$ and s. Assuming the expected marginal utility of effort is independent of s, the analogous result for s follows.

chased, l^*, as

$$(5) \qquad l^* = h(s, \hat{w}^d).$$

The function h—the labor extraction function—summarizes the effects of all of the relevant preferences of the worker, as well as the worker's sense of commitment, injustice, resentment, deference, patriotism, or whatever may affect the difficulty or ease of extracting labor from labor power, or influence the efficacy of surveillance or the threat of income loss as instruments towards this objective.[14]

It is assumed that the employer knows the h function of each worker, and that each is identical, thus allowing one to argue in terms of a representative worker. Further, on the basis of the reasoning above, for both s and $\hat{w}^d$ positive and $l^* < 1$, h_s, and $h_{\hat{w}^d}$ are positive, and $h_{s\hat{w}^d}$ is also positive.[15]

Letting p_x represent a vector of prices of nonlabor inputs, the problem for the employer is now to maximize

$$(6) \qquad R = f(X, L) - p_x X - (w + p_s s) Lp,$$

subject to

$$(7) \qquad L = l^* Lp = h(s, \hat{w}^d) Lp,$$

or to maximize

$$(8) \qquad R = f\left[X, h(s, \hat{w}^d) Lp\right] - p_x X - (w + p_s s) Lp.$$

Because it has been assumed for the moment that the nonlabor inputs X do not affect the labor extraction process, the production function and the extraction function

[14] Before bringing together the three functions—production, cost, and extraction—to consider formally the capitalist's profit-maximizing problem, it may be useful to scrutinize more carefully the nature of the extraction problem. Is this not just another case of the economics of lemons, in which the employer must pay some costs to find out which workers will work hard (or well) and which will not? While some of the results are similar, not all are, and the mechanisms are quite different. The problem for the employer is not to find out what the worker *is*, but to find out what the worker *does*. To see that this is the case, the extreme assumption is made that the employer may know at zero cost the workers' skills and personality characteristics relevant to work motivation and capacities, including exact knowledge of the determinants of the typical (and therefore every) worker's work effort. One of the determinants of work effort is the threat of job loss and hence the level of surveillance. The employer, by these assumptions, knows exactly how much work each worker will do on the average once the employer has selected the level of surveillance and the wage (given external wages, unemployment probabilities and unemployment insurance). At a given moment, however, the employer does not know what the worker is doing, unless the worker is being observed at that moment. And unless the worker is observed not working up to standard, it would not be rational for the employer to fire him or her, for this would convince the remaining workers that the probability of job loss did not depend on work effort, and would thus lower the efficacy of the surveillance inputs. Note that by firing the worker the employer does not eliminate a "bad worker" in favor of a chance at getting a "better worker" from the unemployment pool, for all workers are identical. The purpose of firing the nonworking worker is to convince workers that the surveillance system is effective, and that firing is related to low work effort. In other words, without firings or with firings not based on observed low work effort, the h function would shift adversely from the standpoint of the employer. Strictly speaking, then, the cost of surveillance is not an information cost at all (or at least a very peculiar one) as surveillance will affect increases in effort (over some range) even if the "surveillors" do not pass the information along to the employer, as long as the workers believe that the probability that a nonwork strategy will be detected is a positive function of the level of surveillance. But if employers know exactly how much work each worker will do once the wage and level of surveillance is selected, would it not be optimal to pay workers according to the amount of work done? It might. But this in no way would affect the results below, for the firm's costs will still depend on the number of hours hired (because surveillance s is proportional to hours of labor engaged, not the amount of work done and because workers use inputs owned by the firm). And as long as costs are not independent of the number of hours hired, employers will not be indifferent to how hard each particular worker works. (We will see below that the limiting case of no surveillance inputs cannot be optimal. It is, of course, possible to devise combinations of incentive pay and surveillance such that costs would be independent of hours hired. But it would be quite accidental if that scheme coincided with the optimal incentive structure, given workers' preferences and other relevant information.)

[15] More formally, because the derivative of expected income with respect to work intensity is simply $\hat{w}^d p^d$, the effect of an increase in $\hat{w}^d$ on the workers optimal effort level will depend positively on the level of s, and conversely.

(equations (1) and (5)) are separable, and the employer's maximizing problem may be solved sequentially. The first problem for the employer, and the one that interests us here, is to minimize the cost of a unit of work done, or

$$(9) \qquad \min c_l = (w + p_s s)/h(s, \hat{w}^d).$$

Having solved this problem, its solution, c_l^o, can then be considered the minimum cost of a unit of labor and entered into the employer's new maximand

$$(8') \qquad R = f(X, L) - c_l^o L - p_x X.$$

Assuming, for the moment, an interior solution, and noticing that the marginal cost of a unit increase in $\hat{w}^d$ is one by definition, minimizing (9) requires that

$$(10) \qquad h_{\hat{w}^d} = h(s, \hat{w}^d)/(w + p_s s) = h_s/p_s,$$

or that the average effort per dollar of wage and surveillance cost equal the marginal effort per dollar increase in either wage cost or surveillance cost. Analogously the profit-maximizing employer's strategy must satisfy the condition

$$(10') \qquad p_s = h_s/h_{\hat{w}^d},$$

or the price of surveillance must be equal to the "marginal rate of substitution" between income loss if fired and probability of detection in the labor-extraction function (5).

We may represent this graphically as in Figure 1. The isocost function is a locus of equally costly employer strategies. Because the cost to the employer of a unit increment in $\hat{w}^d$ is one by definition, the slope of the isocost function can be seen to be $-p_s$. The isowork function, derived from the labor extraction function (5), is one of a family of loci of equally effective employer strategies: points describing an equal extraction of labor from a given number of hours of labor power hired. Its slope is $-h_s/h_{\hat{w}^d}$. The expansion path is the locus of all possibly profit-maximizing strategies, namely, those satisfying (10'). Some point on the expansion path, say point a, minimizes the cost of a unit of labor

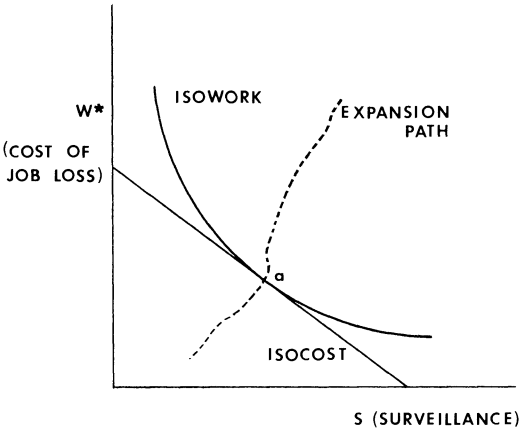

FIGURE 1

and is therefore the solution to (9) and is the profit-maximizing strategy. (It cannot be read directly off the figure.)

I now use this model to demonstrate the three substantive propositions with which I began.

II. The Reserve Army of the Unemployed

The more or less permanent existence of involuntary unemployment is central not only to the Marxian critique of capitalist society, but to the analytical underpinnings of its theory of profit (or of surplus value) as well. Because profits in the Marxian model are not a return to a scarce input but are simply a deduction from total output made possible by capital's power over labor, a complete Marxian model must provide a compelling account of how this power is perpetuated in an economically competitive and politically liberal environment. The basis of this account is the asymmetry between two forms of competition: that among capitalists in selling their outputs and that among workers in seeking employment. Because profit is not the return to a scarce input, in the absence of such an asymmetry, there would be no reason why price competition among capitalists would not drive the profit rate to zero.

The necessary asymmetry is based on the permanent existence of involuntary unemployment, or what Marx termed the reserve

army of the unemployed.[16] The effect of involuntary unemployment is to render labor power nonscarce and hence incapable of claiming the whole product (net of depreciation) through the normal process of competitive price and wage determination. The puzzle is then no longer why profits are not competed away, but why does a nonscarce input, labor power, receive any competitive remuneration at all. The capital-labor distributional conflict thus appears as one taking place between and among two sets of actors, none of which exercise their claims on the product on the basis of a competitively determined return to scarcity in the usual general equilibrium sense.

The Marxian solution to this puzzle is to not reject the competitive assumptions underlying the general equilibrium model, but to pose a distinct theory of the long-term determination of wages and effort in which the former varies negatively and the later positively with the level of unemployment.[17] Only *involuntary* unemployment will affect the bargaining power of capital and labor; hence the centrality of involuntary unemployment to the Marxian theory of the capitalist economy.

On what basis can involuntary unemployment be represented as a general—rather than ephemeral—characteristic of the capitalist economy? The endogenous perpetuation of the reserve army of the unemployed could be assured by a variety of mechanisms: for example, an infinitely elastic supply of labor from other countries or from declining domestic noncapitalist economic systems, such as household production, or rapid structural and technical change accompanied by downwardly sticky wages.[18] The above model of the extraction of labor from labor power points to another possibility, and one more consistent with competitive assumptions, namely, that the labor market does not clear in equilibrium. Put somewhat differently, excess supply in labor markets does not imply a competitive response of wage reductions.

By equilibrium in the labor market, I mean a level of wages, employment, and labor intensity that none of the agents would have both the motivation and the ability to alter. A non-clearing-labor-market equilibrium requires that profit-maximizing employers offer workers a wage and surveillance package such that, given the levels of work effort that workers will choose to expend under the package offered, workers are not indifferent between working and being unemployed. This is, of course, tantamount to saying that a profit-maximizing employer would refuse the offer by a currently unemployed worker to work as hard as the current work force for a wage less than the current wage. We shall see why this counterintuitive result may quite generally occur.

It is clear then that a market-clearing wage would imply that in our model the cost of job loss is zero, for if the cost is not zero the worker cannot be indifferent between employment and unemployment. Under what conditions could a wage-surveillance package that rendered the worker indifferent between employment and nonemployment be a profit-maximizing strategy for the individual employer, and hence a possible equilibrium? Or, in terms of Figure 1, could an optimal strategy lie on our horizontal axis, indicating a zero income loss associated with being fired? Because $h_s \to 0$ as $\hat{w}^d \to 0$, and analogously $h_{\hat{w}^d} \to 0$ as $s \to 0$, the expansion path for any $p_s > 0$ will lie entirely within the range of positive values of s and $\hat{w}^d$. As long as the employer has hired some surveillance inputs, a market-clearing wage ($\hat{w}^d = 0$) can-

[16] Marx (1976): "...relative surplus population (i.e., unemployment, SB) is therefore the background against which the demand and supply of labor does its work" (p. 792). And, "The pressure of unemployment compels those who are employed to furnish more labor and therefore makes the supply of labor to a certain extent independent of the supply of workers. The movement of the law of supply and demand on this basis completes the domination of capital" (p. 793).

[17] The macroeconomic and general equilibrium characteristics of this solution are the subject of two of my other papers (1983a,b).

[18] If the supply of labor hours is infinitely elastic at a given wage, those who are not employed cannot be said to be involuntarily unemployed strictly speaking, as they are unwilling to offer any labor time at a lower wage.

not be optimal (because $h_s = 0$ for $\hat{w}^d = 0$). The critical role of the cost of surveillance is here clearly indicated, for with $p_s = 0$ the isocost functions in Figure 1 would be horizontal: (free) surveillance would be substituted for (costly) job loss threat and the cost minimum would occur at $\hat{w}^d = 0$, a result consistent with the traditional market-clearing equilibrium.

But what of the "no income loss, no surveillance strategy" represented by the origin in Figure 1? In order for this strategy to be optimal, it would have to be the case that

$$(11) \quad h(0,0)/w > h(s,\hat{w}^d)/(w + p_s s)$$

for all possible levels of s and w. In this case, surveillance and job loss threats are sufficiently ineffective or costly to prohibit their use at any level. But this implies that, even when it is possible for the employer to exercise power over the worker, it is not profitable to do so. But this could only be true if there were no conflict of interest between the worker and the employer. In this case, employer and workers have a "conflict of interest" only in the socially irrelevant sense that sunbathers and drought-stricken farmers have a conflict of interest (barring the possibility of rainmaking).

This result does not depend on the manner in which the probability of reemployment (j) is determined. Assume for the moment that the government committed itself to achieving full employment, either through fiscal and monetary policy, or simply by guaranteeing any unemployed worker a job at the going wage. With $j = 1$ the employer might either set $\hat{w}^d > 0$ by offering a wage higher than other employers, or set $\hat{w}^d = 0$. The former is inconsistent with equilibrium. This can be readily seen by rewriting the cost of being fired as an equilibrium condition (with $w = \hat{w}$) or $\hat{w}^d = (1 - j)(w - w^c)$. By the logic of the previous paragraph, the latter is inconsistent with the assumed conflict of interest between worker and capitalist.

Let us summarize these results. Given a positive cost of surveillance and a conflict of interest between employer and worker over work effort, the wage rate offered by the competitive profit-maximizing employer will exceed the worker's next best alternative. This is possible in general only if the probability of reemployment is less than one. Therefore, labor market competition cannot clear the labor market. Correspondingly, market clearing—the absence of involuntary unemployment—implies labor market disequilibrium.[19]

Other than ruling out market clearing as a possible labor market equilibrium, this model bears no direct implications concerning the determination of the general level of unemployment or the probability of reemployment. But it does provide a microeconomic foundation consistent with Kalecki's suggestion that sustained full employment and the long-run survival of capitalist enterprise may be inconsistent. Indeed, given a conflict of interest between employer and worker, labor market clearing implies either escalating wage increases, or a reduction in work effort to those levels chosen by workers. Particularly in an open economy, neither result would likely be conducive to investment levels capable of sustaining full employment (but to pursue this argument we would have to go considerably beyond the microeconomic confines of this paper).[20]

These results would be modified, of course, if employers were assumed to have not pro-

[19] This result is similar to that produced—with somewhat different models—by Calvo, B. Curtis Eaton and William White (1982), James Malcomson (1981), Hajime Miyazaki (1981), Tekashi Negishi (1979), Solow (1980), Carl Shapiro and Joseph Stiglitz (1984) and others. In all of the above, actual amount of work done is directly or indirectly a positive function of the wage rate. Miyazaki focuses on the problem of worker free riding against other workers in a work group. Eaton and White focus on "trust jobs." Malcomson assumes "at least two types of individuals with different productivities who cannot be discriminated perfectly by observation at work" (p. 865). Negishi and Solow both base their models on problems of worker morale and "affront" (Negishi, p. 114). Closest in spirit to my model (though lacking the surveillance element) is Calvo, who, however, while demonstrating the possibility of nonclearing equilibria, assumes an interior solution to a problem analogous to the minimization of (6), thus eliminating the market-clearing equilibrium by assumption.

[20] I develop this argument in my 1981, 1983a,b papers. See also Gintis and Tsuneo Ishikawa (1983).

hibitively expensive ways of imposing effective sanctions on workers even in the absence of involuntary unemployment. The extent to which such alternative sanctions are feasible and effective is in part an empirical issue that cannot be resolved here. For whatever reason, the practical import of most of the alternatives to the threat of involuntary unemployment appears to be quite limited in the U.S. economy.[21]

[21] If workers could instantaneously find alternative employment, but nonetheless bore significant costs of job changing—either through moving costs, training costs not borne by their new employer, employment bonds, or job entry fees that are forfeited upon job loss, a tax levied by the government on job changers, or through any other means, or if on-the-job nonwork activities were treated as a criminal offense subject to fines or imprisonment, the attainment of full employment could not be ruled out on theoretical grounds. While possible substitutes (or complements) to the threat of unemployment are thus readily imaginable, their actual or potential relevance to the problem of getting workers to work may be questioned. First, to replace the threat of unemployment, the costs imposed must be quite substantial, considerably more than reasonable moving or training costs, and in excess of what most workers can readily borrow for payment of an employment bond. Juliet Schor and I (1983) estimate that in 1983, for example, the mean cost of job loss (roughly an after-tax estimate of $\hat{w}^d$) was about one-half the mean after-tax annual income of a fully employed production worker. (This is a low estimate, as it abstracts from the costs associated with the loss of job seniority.) Moreover, the variance among individuals of the expected cost of job loss is probably quite large, due to the high variance of unemployment duration, suggesting that if we were to drop the unrealistic assumption that workers are risk neutral, the certainty equivalent of the cost of job loss might be considerably greater than Schor's and my estimates. Consideration of the social or psychological costs of unemployment—even with a generous accounting of the joys of free time—would further augment the estimate of the costs of job loss. Second, the imposition of these alternative sanctions by either employers or through the government may involve private or social enforcement costs, or other welfare losses sufficiently large to inhibit their use. Third, some otherwise promising methods of eliciting work effort other than the threat of unemployment may be considered to be socially unacceptable or politically infeasible. Even assuming that effective alternative sanctions were feasible would only modify rather than nullify my results unless these alternatives were so cost effective as to totally eclipse the expedient of paying workers more than their supply price.

III. Capitalist Technology

Central to the Marxian critique of capitalist society is the idea that the competitive pursuit of profits requires employers to organize the production process so as to maintain their power over workers, and that at least some of the boredom, fragmentation, and other undesirable aspects of the work experience may be attributed to this fact and not to the requirements of technical rationality. According to this view, the prevailing organization of production—including the technologies in use—cannot be derived solely from an interaction of exogenously given technical possibilities and worker and consumer preferences for goods, leisure, and various kinds of work environment, but rather reflect the class interest of capital as well. Hence the expression "capitalist technology."

To suggest that technology may be an instrument of class conflict does not mean, of course, that employers may select technologies without regard to the competitive requirements of cost minimization. Nor does it require that capitalists collude in their choice of production methods or in the development of future technologies. Rather, the concept of capitalist technology is based on the proposition that cost minimization by competitive employers implies the selection of profitable but inefficient technologies even in the absence of market failures arising from collusion, externalities, extended time horizons, and the like.

I will say that the capitalist has chosen an inefficient technology when there exists some other method of production that, per unit of output, uses less of at least some input and not more of any. The logic of the concept of capitalist technology is that a technology that is inefficient in the above sense may nonetheless be cost minimizing if it allows the capitalist to lower the cost of some input. This is possible in the Marxian model because the firm is not a price taker with respect to the price of labor, but rather may alter this cost through the selection of various labor extraction strategies. The most obvious case of this is the adoption of machine-paced production as a means of

increasing the intensity of labor.[22] In this case, costs may be lowered not only by producing more with the same inputs, but by extracting more of one of the inputs—labor—for the same price, and thus lowering the unit cost of labor. Machine-paced production may of course also be efficient. But it is simple to show that it *need not* be efficient in order to be adopted.

Capital goods may be considered to be capable of joint production, simultaneously contributing to the marketed output of the firm and producing or contributing to the acquisition of information on the work performance of the workforce. The assembly line, and even factory production itself (in contradistinction to more decentralized production methods), as well as modern information-processing systems are important cases of surveillance information-producing technologies.

The implications for efficient technical choice may be readily seen by modifying the labor extraction function to take account of this form of joint production. We now have

$$(5') \qquad l^* = h\left[p^o(s, x), \hat{w}^d\right],$$

where x is the vector of inputs (per labor hour) of production equipment and intermediate goods, and $p^o(s, x)$ is the worker's expected probability that a nonwork strategy will be detected. For some x we have $p_{ox} > 0$ and hence $h_x > 0$: given the cost of job loss ($\hat{w}^d$) and the level of (pure) surveillance inputs (s), the use of larger amounts of some input in the production process will increase the amount of work done per hour by increasing the probability that a nonwork strategy will be detected, thus increasing the worker's expected cost of pursuing a nonwork strategy.

It can be seen in this case that even if all relative goods prices were optimal (in the sense that they accurately reflected relative scarcities), the familiar conditions for efficient technical choice (i.e., $f_x = p_x$) would be violated. For it will now be the case that the profit-maximizing employer will maximize profits by observing the following condition:

$$(12) \qquad f_x + f_{l^*} h_x = p_x.$$

The second term on the left-hand side reflects the contribution of a marginal increment in x to production via its contribution to the extraction of labor from labor power. (It is redundant to observe that under these conditions the relative general equilibrium prices would also not be optimal.)

The implication of this point is that a competitive profit-maximizing capitalist could choose a technology using more of both x and l^* per unit of output. This may be readily seen by noting that the isocost function slope is

$$(13) \qquad dl^*/dx = -(p_x + l^* c_{lx})/c_l,$$

where c_{lx}, the derivative of the cost of a unit of effort with respect to x, is negative, and hence the numerator is not necessarily negative. Thus the isocost function may be positively sloped, leading to the possibility that cost minimization may result in the choice of an inefficient technology, namely in the rejection of a technology using less of both l^* and x per unit of output.

It might be thought that this demonstration implies that the need for surveillance inputs is somehow illegitimate and should be abstracted from in consideration of efficiency. Indeed, as we shall see in the penultimate section, the assertion that the class structure of capitalism induces a particularly high level of work resistance and hence promotes the extensive use of surveillance inputs differentiates the Marxian from the neo-Hobbesian view. But the above argument involves neither abstracting from surveillance inputs, nor considering surveillance to be a kind of false need induced through an endogenously generated disutility of labor.

Quite the contrary, pure surveillance inputs s, with an *exogenously* determined labor extraction function, provide a particularly clear case of the above argument. Consider the indicated isowork function in Figure 1 as representing an amount of work effort capable of producing one unit of output. Starting at point a, were the firm to move along this

[22] Edwards (1979) refers to this as "technical control" in contradistinction to "bureaucratic control" or "simple control" of the production process.

isowork locus by raising wages and cutting surveillance inputs, the cost of labor would rise and hence the profit rate would fall, but output per unit of input would rise (l^* remaining constant and s falling). This result arises because there is a tradeoff between surveillance and the wage rate in the labor extraction function, and while surveillance inputs are resource-using, the wage rate is not; hence raising wages and lowering surveillance may be efficient but not profitable. Thus cost minimization and efficiency do not coincide: the tradeoff in this case is not efficiency vs. equity, but efficiency vs. profitability.

IV. Divide and Rule

Central to recent Marxian research on racial and sexual discrimination, segmented labor markets, and internal labor markets is the proposition that divisions among workers may be in the interest of employers, and further that it may be in the interest of competitive noncolluding employers to discriminate among workers on the basis of ascriptive characteristics unrelated to the individual worker's ability or willingness to contribute to the production process.[23]

Reich, Roemer, Gintis, and others have recently proposed coherent models of discriminating competitive capitalists. The present model of the extraction of labor from labor power may be extended in a very simple way to capture the logic of these contributions.[24]

We say that an employer discriminates when he or she makes different wage-surveillance offers to workers of differing ascriptive characters (race, sex, age) who are otherwise identical with respect to their productive capacities and proclivities, that is, given that we have assumed that labor services are homogeneous, identical with respect to their labor extraction functions, h. I now introduce the possibility that workers may cooperate either to render surveillance more difficult or otherwise more expensive (for example by refusing to offer information on the work or nonwork activities of fellow workers), or to reduce or withdraw labor services should the employer treat a fellow worker in a manner thought to be unjust or simply contrary to the interests of other workers. Labor services may be withdrawn either through a reduction in work effort (an outward shift in the h function), or in an extreme case through a strike.

The extent of worker cooperation, including the possibility of forming institutions such as unions, varies positively with the extent of worker unity, u. Worker unity will depend on general social conditions external to the firm, but it will also be influenced by the firm's hiring and pay policies. Where a uniform wage surveillance package is offered to all workers, for example, opportunities for joint negotiations concerning wage and working conditions will be enhanced, and divisive sentiments such as envy and invidious distinction attenuated. With distinct pay and surveillance packages offered to different workers—particularly to groups of workers predominantly composed of individuals of different race, sex, age, and other characteristics—employers are more likely to be able to bargain separately with each group to foster competition, envy, or even hostility among the distinct groups, and thus to discourage unity. For simplicity we say that unity, u, will be a negative function of a measure of wage inequality of the workforce of the firm, v.[25]

[23] This view may be distinguished from that which maintains that the cost-minimizing process renders discrimination unprofitable to the individual employer, however beneficial it might be to the employer's class as a whole, and hence that discrimination is primarily an ideological or political phenomenon whose perpetuation is explained by inertia, ignorance, or by the collective action (in the media, schools, state, or elsewhere) of those who benefit from it.

[24] See Reich and the previously cited references to Roemer and Gintis. This model differs somewhat from those cited in stressing the costly nature of surveillance and the cost of job loss rather than bargaining strength based on worker unity. All of the Marxian models differ from the search theory approach to the stability of discrimination in a competitive environment in that the employer is assumed to know all of the relevant worker characteristics.

[25] Because u cannot readily be measured, this behavioral assumption cannot easily be tested. But it is strongly supported by the relevant works in labor economic and labor history. See Reich, and Edwards, Gordon, and Reich and the works cited therein.

I make an additional assumption, not necessary to my result but one which will enrich the model somewhat: let us now assume that there are some costs to the employer of replacing the worker (firm-specific training, or other), and that, for this reason, when a worker is detected pursuing a nonwork strategy, the employer may choose not to terminate the worker's employment.

In my expanded model, then, the expected cost to the worker of pursing a nonwork strategy is

$$E(n) = p^o p^t \hat{w}^d, \tag{14}$$

where, as before $\hat{w}^d$ is the cost of job loss, p^o is the probability of being detected should the worker pursue a nonwork activity, and p^t, previously assumed to be unity, is now the variable probability of being terminated, if detected. By the above argument,

$$p^o = p^o(s, x, u) \quad \text{with } p_{ou} < 0;$$

$$p^t = p^t(u) \quad \text{with } p_{tu} < 0.$$

The labor extraction function thus becomes

$$l^* = h\left[p^o(s, x, u), p^t(u), \hat{w}^d\right] \tag{5''}$$

in which the derivative of l^* with respect to u is negative, taking account of the effects of unity on both the probability of detection and the probability of termination.

Under what conditions will the employer described in this model choose to discriminate? Assume that there are two "types" of worker, type i and type j. Why would the employer pay them different wages? It is clear at once that if the wage rates prevailing in the rest of the economy are different, or if the probability of reemployment or access to unemployment insurance is different, the optimal wage offers w_i and w_j will differ. Thus given differing external conditions, the firm will choose to offer differing wages to each type of worker. But it will be clear that the cost of a unit labor from one type of worker is less than the other, or $c_{lj} < c_{li}$ (assuming that type i workers are favored by higher wages and/or reemployment probabilities or access to unemployment insurance in the remainder of the economy). So the question arises, why would the employer choose to employ any of type i?

Assume that the employer hired no type i workers. In this case, there would be no wage inequality among the workforce ($v = o$). Hiring some type i workers will yield a positive v, thus increasing l^* and possibly lowering the average cost of labor for the firm as a whole, c_l. By the same reasoning, it could be in the interest of the employer to offer type i and type j workers different wages, even if in the rest of the economy they were treated perfectly equally. Moreover, given the existence of involuntary unemployment, such a strategy would not be rendered infeasible by the labor supply choices of the group which was offered the lower wage.

A related but distinct argument for paying identical workers different wages may also be offered, if the model is extended to more than one time period. Assume initially that all workers are paid the same wage. An employer could then offer a prospective worker a two-period wage package with a low first-period and high second-period wage. The difference in the first-period wages under the equal wage and the stepped-wage package may be considered an employment bond paid by the worker to the employer which will be returned to the worker in the form of higher second-period wages, unless, of course, the worker is fired in the interim. Let the wage cost to the firm of the two packages be the same, assuming the firm intends to make good its second-period offer, and expects the worker to neither quit, nor be fired. The "less now, more later" offer will elicit more work from the worker, however, because once it is accepted and work under its terms has commenced, the cost of job loss under the terms of that package is greater, because the worker has already performed some low wage labor and has an increasingly advantageous balance of high wage labor to look forward to should he or she retain the job. In a regime of generalized stepped-wage offers such as the primary labor market in the United States, the costs of failing to cash in on later-period high wages can be considerable.

The worker may not accept the stepped-wage offer, of course, if he or she believes that the probability of getting arbitrarily fired

at the end of the first period is high. But should the worker accept the stepped-wage offer, the firm will have affected a reduction in its cost of labor c_l. As in the case of discrimination above, the fact that jobs are rationed will allow the firm to recruit labor using the less attractive stepped-wage offer.

Thus long-term contracts and internal labor markets—promotion ladders according to job tenure and unrelated to skill—may be a method of increasing the cost of job loss to the worker without increasing the wage bill, and hence an effective means of reducing the cost of labor (in effort units).[26]

The above explains why identical workers may be paid differently. It does not explain why discrimination exists, or why type *i* workers tend to be white, male, and neither very young nor very old. But it does present one possible argument for the reproducibility of discrimination and internal labor markets in a competitive capitalist economy.

V. Neo-Hobbesian and Marxian Models

It may well be objected that while the labor extraction model provides an internally consistent analysis of involuntary unemployment, inefficient technical choice, and discrimination in a competitive equilibrium, any negative normative connotations would be misplaced, for these undesirable outcomes might be intrinsic to *any* system of production, irrespective of the social structure in which it is embedded. Indeed this is precisely the implication of what I have termed the neo-Hobbesian models of the production process.

Malfeasance is to the neo-Hobbesian models what class conflict is to Marxian models. The key difference between the two is this: malfeasance is a universal human proclivity —in this case based on the inherent nature of work as a disutility. By contrast, class conflict in the labor process of a capitalist economy is the result of a specific and mutable set of social institutions; the conflict over work intensity being *at least in part* the consequence of the particular organization of work and the resulting alienated nature of labor.

Samuelson's statement cited at the outset —while based on a Walrasian model—reflects the spirit of the neo-Hobbesian models as well, for it is consistent with the view that the form of the class relationship imparts nothing of importance to the production process.

Can the neo-Hobbesian position be sustained? Can the Marxian problem—class conflict over the extraction of labor from labor power—be reduced to the general problem of malfeasance? Differing ideological connotations aside, is the extraction of labor from labor power simply another way of addressing the universal problem of "shirking"?

Concern with the general problem of reconciling individual self-interest and collective rationality is hardly new, dating back at least to Hobbes. That the regulation of self-interest through the market provided a solution to the Hobbesian problem was suggested metaphorically by Mandeville during the eighteenth century and developed fully by Walras and by twentieth-century welfare economists. If all economic interactions are contractual exchanges, the conflict of self-interest and collective rationality is capable of resolution, or at least substantial attenuation.

But, as economists of all persuasions now recognize, not all economic interactions are exchanges. Coase's conception of the firm, as a command economy of nonexchange relations, is a necessary but possibly troublesome addition to any analysis of a specifically capitalist economy characterized by an

[26] From quite different perspectives, a similar argument has been suggested by Edwards and by Lazear. The argument is quite distinct, however, from models based on search theory and screening costs, in which the employer has an interest in retaining the worker (because of hiring costs). See, for example, Okun. The post-World War II emergence of long-term contracts and internal labor markets as characteristic of a major segment of the U.S. economy may be attributable in part to their labor extraction cost-saving aspect, to the historically low rates of unemployment in the postwar period, and to the apparent decline in the cost of job loss associated with a spell of unemployment. Further, as Lazear has pointed out, the labor extraction advantages of long-term stepped-wage offers may help explain the otherwise anomalous phenomenon of returns to job tenure significantly in excess of any empirically compelling estimates of productivity enhancement through generalized on-the-job learning. See James Medoff and Katherine Abraham (1980).

employment relation. Strikingly, the Coasian view of the capitalist economy as a multiplicity of mini-command economies operating in a sea of market exchanges is radically different from the Walrasian foundations of welfare economics, and superficially indistinguishable from the Marxian view.

The question obviously arises, then, as to the compatibility of the Coasian insight (command) and the Mandevillian solution to the Hobbesian problem (markets). Are the command relations of the firm a rational solution to the problem of the coordination of individual and group rationality? Or are they, in some sense, a market failure attributable to the successful pursuit of the interests of those who command the firm? This is the central issue dividing the neo-Hobbesian from the Marxian analysis.

Coase, basing his concept of the firm on the notion that command relations supercede market relations when the transactions costs of markets exceed the analogous costs of command and nonmarket coordination, initiated a literature which affirmed the efficiency of the hierarchical structure of the firm.[27] Because malfeasance is no more than an expression of the natural self-interestedness of human beings, the cost of policing malfeasance cannot be considered evidence of a failure of markets. The logic of this position can be illustrated within the terms of the Marxian model.

Let us make the (neo-Hobbesian) assumption that the labor extraction function is given by human nature. People's attitude towards work—broadly, the disutility of labor—is unrelated to the social institutions that govern the process of work. In this case, the extraction function must be considered to be exogenous, not only to the firm but to the society as a whole. Hence the various employer strategies and their results must be considered to be little more than a consequence of the (possibly lamentable but ineradicable) human tendency to avoid work. A society might nonetheless choose to discourage discrimination, to minimize involuntary unemployment, or to discourage the use of surveillance equipment or personnel, but they would do so only at the cost of choosing to permit a higher level of what the neo-Hobbesian literature terms free riding or shirking, and consequently a lower average level of output per hour of labor.

But the assumptions required to sustain the neo-Hobbesian view are exceptionally restrictive and implausible. We have seen in the analysis of capitalist technology that even with an exogenously given labor extraction function, the choice of technology—including the level of surveillance—which is profit maximizing will not in general be efficient: it generally will be dominated by some other less profitable and less surveillance-intensive combination of inputs.[28]

Perhaps more fundamentally, the assumption of an exogenous extraction function appears to be quite arbitrary. If the organization of the work process and the principles determining the distribution of the net revenues arising therefrom influence workers' attitudes towards work and hence are among the determinants of the extraction function, the neo-Hobbesian conclusions are considerably altered. In this case, there may exist some alternative set of arrangements in which a bargain could be struck in which at least one of the participants was better off and none worse off. A possible argument may be illustrated. Rewrite the labor extraction func-

[27] The recent literature was initiated by Alchian and Demsetz.

[28] Because the efficient (less surveillance-intensive) technology is less profitable it might be objected that while the neo-Hobbesian position is faulty on static efficiency grounds, a dynamic efficiency perspective, taking account of optimal levels of investment and the relationship of profits to investment, would salvage their view. But this is not the case unless it is also assumed that the current levels of investment are at or below the optimal level and further (and dubiously) that a reduction in the profit rate is necessarily associated with a decline in investment. To the extent that capitalists consume rather than invest their profits (or invest them in other economies), a decline in the profit rate does not *require* a reduction in the level of investment, even if the economy is operating at the level of potential output. Of course, given the institutions that define the capitalist economy, such an effect is likely to result, but it is hardly reasonable to take as given the institutions which are themselves under evaluation.

tion as

$$(5''') \qquad l^* = h(i, s, \hat{w}^d, u, x),$$

where i is a vector reflecting the general institutional environment. If it could be shown that in an environment which workers perceived to be more fair, or more consistent with their self-respect, for example, they would choose to expend more effort for any given employer strategy, then it is a simple matter to demonstrate that the initial outputs could be produced with unchanged levels of labor effort in production and using less surveillance labor.[29] In Figure 1, the transformed institutional environment (the change in i) would be reflected in an inward shift in the isowork loci such that the initial amount of work could be extracted with a reduced s.[30] The newly released surveillance labor could then be employed producing goods representing a net addition to the total product, achieved without increasing total labor hours worked and/or workers' efforts per hour.

The above argument draws directly on the third basic characteristic of the production process in the Marxian model, the joint production of commodities and workers or the endogenous nature of workers' preferences. The attitude towards work is not, according to this principle, simply a manifestation of human nature, but in part the result of the social institutions in which the production process takes place.

In the production of workers, of course, other institutions—schools, the family, political organizations, and the like—assume a critical importance. The structure of these institutions is, however, strongly albeit indirectly influenced by the structure of the production process.[31] Moreover, the structure of the production process itself undoubtedly has direct effects on attitudes towards work. A more democratic structure of decision making and a more egalitarian distribution of the firm's net revenues, for example, might both reduce the incentive to pursue nonwork activities and heighten the cost of so doing by enlisting fellow workers as more ardent enforcers of the pace of work, or more willing cooperators with the surveillance system.[32]

The neo-Hobbesian's normative position thus seems dubious on two grounds: the discrepancy between profitability and efficiency, and the endogeneity of the labor extraction function. If the social nature of the labor extraction function is conceded and, further, if the feasibility of forms of social structure and work organization conducive of lower levels of work resistance or higher levels of work motivation is accepted,

[29] There seems to be considerable evidence that this is the case. See, for example, Raymond Katzell et al. (1975).

[30] A simple reduction of s would not be optimal, of course, but this is immaterial to my argument.

[31] The influence is mutual, of course, schools and families influencing the structure of production as well as conversely. See, from very different perspectives, my book with Gintis (1976), Melvin Kohn (1969), and William Lazonick (1978, 1981). Lazonick concluded, "Hence it can be argued that not only the institutional transformation of the capitalist enterprise but also, and perhaps more fundamentally, the institutional transformation of the larger society was required to stabilize the capital labor relation in the mass production industries" (1981, p. 36).

[32] Why are the potential gains to such an alternative form of work organization not sufficient to bring such worker-based enterprises into being and to assure their success in the competitive struggle with more hierarchically structured capitalist firms? If workers' attitudes toward work were determined solely and instantaneously by the work environment in which they worked, and if credit were readily available on terms no worse than those available to capitalist firms, any group of workers could form a co-op and reap the benefits of lessened surveillance. Both assumptions are highly questionable. To the extent that attitudes toward work are determined by an entire nexus of social institutions which change slowly, the opportunities for the atomistic movement towards a less socially irrational form of production are quite limited. Perhaps more important, because workers' own assets are not extensive, their access to credit is limited or costly by comparison to that enjoyed by the owners of firms. (It matters little for the issues treated here whether the different terms of credit available to capitalists and workers reflect rational profit-maximizing behavior by lenders or an imperfection in the credit market.) And it might be added that, perhaps for some of the reasons outlined in this paper, and despite the obstacles outlined in this note, the last decade has witnessed a substantial growth of workers' co-ops and worker-managed firms in the United States.

or, if the possible nonoptimality of the competitively determined profit rate is admitted, the command relationships within the firm and the associated patterns of involuntary unemployment, technical choice, and discrimination must be viewed as market failures rather than simply as unavoidable transactions costs. Moreover, because of the importance of the labor input in the production process, the quantitative importance of this source of market failure may overshadow the more commonly recognized environmental and other externalities.

VI. Conclusion

The model of the production process based on the extraction of labor from labor power thus provides an internally consistent microeconomic theory capable of supporting some of the most fundamental general propositions in Marxian economics concerning the reserve army of the unemployed, the determination of the profit rate, discrimination, and the irrationality of the organization of work and technology. The above arguments do not, of course, establish the superiority of the Marxian model. Nor do they provide any indication that the Marxian model is capable of generating plausible empirical accounts of such phenomena as movements in the unemployment rate, the profit rate, the structure of discrimination, or technical choice.

However, a significant amount of empirical work along the lines outlined above has been done, some of it with quite successful results. For example, econometric models of postwar U.S. productivity growth, the profit rate, Tobin's Q, and strike activity using an empirical measure of the cost of job loss, $\hat{w}^d$, have generated highly significant and robust estimates consistent with the expectations of this model.[33] Historical studies of technical choice and work organization based on the extraction of labor from labor power have produced compelling accounts of otherwise anomalous patterns of technical change. (See Lazonick, 1982, and Marglin.) And econometric studies of the distributional impact of discrimination have produced results quite consistent with the divide and rule interpretation. (See Reich.) None of these is alone decisive, but taken together they do suggest that the Marxian model offers a promising direction for empirical investigation.

[33] See Thomas Weisskopf et al. (1983), my paper with Gordon and Weisskopf (1983), and Schor's and my papers (1983, 1984). Michele Naples (1982) has estimated significant relationships between labor productivity and the structure of control of the labor process consistent with the above model.

REFERENCES

Akerlof, George, "A Theory of Social Custom, of Which Unemployment May Be One Consequence," *Quarterly Journal of Economics*, June 1980, *94*, 749–75.

Alchian, Armen and Demsetz, Harold, "Production, Information Costs and Economic Organization," *American Economic Review*, December 1972, *52*, 777–95.

Becker, Gary, "Investment in Human Capital," *Journal of Political Economy*, October 1962, *70*, 9–49.

Boddy, Raford and Crotty, James, "Class Conflict and Macro Policy: The Political Business Cycle," *Review of Radical Political Economics*, Spring 1975, *7*, 1–19.

Bowles, Samuel, "Competitive Wage Determination and Involuntary Unemployment," mimeo., 1981.

_____, (1983a) "Long Term Growth and Equilibrium Unemployment in an Open Competitive Capitalist Economy," mimeo., 1983.

_____, (1983b) "The Cyclical Movement of Real Wages, Labor Productivity, and 'Overhead Labor' in a Competitive Non-Monetary Economy," mimeo., 1983.

_____, **and Gintis, Herbert,** *Schooling in Capitalist America*, New York: Basic Books, 1976.

_____, **Gordon, David M. and Weisskopf, Thomas,** "The Profit Rate in the Post-War U.S. Economy: An Econometric Investigation," mimeo., 1983.

Braverman, Harry, *Labor and Monopoly Capital*, New York: Monthly Review Press, 1974.

Calvo, Guillermo, "Quasi-Walrasian Theories

of Unemployment," *American Economic Review Proceedings*, May 1979, *69*, 102–07.

Coase, Ronald, "The Nature of the Firm," *Economica*, November 1937, *4*, 387–405.

Commons, J. R., *History of Labor in the United States*, New York: Macmillan, Vols. 1, 2, 1918; Vols. 3, 4, 1935.

Eaton, B. Curtis and White, William D., "Agent Compensation and the Limits of Bonding," *Economic Inquiry*, July 1982, *20*, 330–43.

Edwards, Richard C., *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*, New York: Basic Books, 1979.

_____, **Gordon, David M. and Reich, Michael,** *Segmented Work, Divided Workers; The Historical Transformation of Labor in the United States*, Cambridge: Cambridge University Press 1982.

Gintis, Herbert, "The Nature of the Labor Exchange," *Review of Radical Political Economics*, Summer 1976, *8*, 36–54.

_____ **and Ishikawa, Tsuneo,** "Wages, Work Discipline, and Macroeconomic Equilibrium," mimeo., 1983.

Glyn, Andrew and Sutcliffe, Robert, *British Capitalism, Workers, and the Profit Squeeze*, London: Penguin, 1972.

Goodwin, Richard, "A Growth Cycle," in C. H. Feinstein, ed., *Capitalism and Economic Growth*, Cambridge: Cambridge University Press, 1967.

Kalecki, Michel, "Political Aspects of Full Employment," *Political Quarterly*, October-December 1943, *14*, 322–30.

Katzell et al., R. A., *Work, Productivity and Job Satisfaction: An Evaluation of Policy Related Research*, New York: Harcourt Brace Jovanovich, 1975.

Kohn, Melvin, *Class and Conformity, A Study in Values*, Homewood: Dorsey Press, 1969.

Lange, Oskar, "Marxian Economics and Modern Economic Theory," *Review of Economic Studies*, June 1935, *2*, 189–201.

Lazear, Edward, "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, September 1981, *71*, 606–20.

Lazonick, William, "The Subjugation of Labor to Capital: The Rise of the Capitalist System," *Review of Radical Political Economics*, Spring 1978, *10*, 1–31.

_____, "Technological Change and the Control of Work: A Perspective on the Development of Capital Labor Relations in U.S. Mass Production Industries," Discussion Paper No. 821, Harvard Institute of Economic Research, 1981.

_____, "Production, Productivity and Development, Theoretical Implications of Some Historical Research," Discussion Paper No. 876, Harvard Institute of Economic Research, 1982.

Malcomson, James, "Unemployment and the Efficiency Wage Hypothesis," *Economic Journal*, December 1981, *91*, 848–66.

Marglin, Stephen, "What Do Bosses Do? The Origins and Function of Hierarchy in Capitalist Production," *Review of Radical Political Economy*, Spring 1974, *6*, 60–112.

Marx, Karl, *Capital*, Vol. I, Harmondsworth: Penguin, 1976.

_____, *Capital*, Vol. III, New York: International Publishers, 1967.

_____, *The Grundrisse: Introduction to Critique of Political Economy*, New York: Vintage, 1973.

Medoff, James and Abraham, Katherine, "Experience, Performance, and Earnings," *Quarterly Journal of Economics*, December 1980, *95*, 703–36.

Miyazaki, Hajime, "Work Norm and Involuntary Unemployment," Discussion Paper No. 7, Stanford Workshop on Factor Markets, July 1981.

Naples, Michele, "The Structure of Industrial Relations, Labor Militance and the Rate of Growth of Productivity: The Case of U.S. Mining and Manufacturing, 1953–1977," unpublished doctoral dissertation, University of Massachusetts, 1982.

Negishi, Tekashi, *Microeconomic Foundations of Keynesian Macroeconomics*, Amsterdam: North-Holland, 1979.

Okun, Arthur, *Prices and Quantities, A Macroeconomic Analysis*. Washington: The Brookings Institution, 1981.

Reich, Michael, *Racial Inequality*, Princeton: Princeton University Press, 1980.

Roemer, John, "Divide and Conquer: Micro Foundations of the Marxian Theory of Discrimination," *Bell Journal of Economics*, Autumn 1979, *10*, 695–705.

Samuelson, Paul, "Wage and Interest: A Mod-

ern Dissection of Marxian Economic Models," *American Economic Review*, December 1957, *47*, 884–912.

Schor, Juliet and Bowles, Samuel, "Conflict in the Employment Relation and the Cost of Job Loss," mimeo., 1983.

_____ **and** _____, "Employment Rents and Class Conflict: An Empirical Investigation," mimeo., 1984.

Shapiro, Carl and Stiglitz, Joseph, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, June 1984, *74*, 433–44.

Solow, Robert, "Alternative Approaches to Macroeconomic Theory: A Partial View," *Canadian Journal of Economics*, August 1979, *12*, 339–54.

_____, "On Theories of Unemployment," *American Economic Review*, March 1980, *70*, 1–10.

Weisskopf, Thomas, Gordon, David M. and Bowles, Samuel, "Hearts and Minds: A Social Model of U.S. Productivity Growth," *Brookings Papers on Economic Activity*, 2:1983, 381–450.

Williamson, Oliver, "The Organization of Work," *Journal of Economic Behavior and Organization*, March 1980, *1*, 5–38.