

NBER WORKING PAPER SERIES

UTILITARIANISM AND HORIZONTAL EQUITY:
THE CASE FOR RANDOM TAXATION

Joseph E. Stiglitz

Working Paper No. 694

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

June 1981

The research reported here is part of the NBER's research program in Taxation. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Utilitarianism and Horizontal Equity:
The Case for Random Taxation

ABSTRACT

This paper establishes that, far from being able to derive the principle of horizontal equity from utilitarianism, the principle is actually inconsistent with utilitarianism in a variety of circumstances. We derive conditions under which (a) it is optimal to impose random tax schedules (ex post randomization); and (b) it is optimal to randomize the tax schedules imposed on a set of otherwise identical individuals (ex ante randomization). The implications for optimal tax theory are discussed. More generally, it is shown that there are a number of potentially important economic situations with which the principle of horizontal equity may be inconsistent not only with utilitarianism but even with Pareto optimality.

Joseph E. Stiglitz
Department of Economics
Dickinson Hall
Princeton University
Princeton, New Jersey 08544

(609) 452-4014

UTILITARIANISM AND HORIZONTAL EQUITY:

THE CASE FOR RANDOM TAXATION*

by

Joseph E. Stiglitz

1. Introduction

The concept of horizontal equity has long had a special place in public finance. In particular, the imposition of a random tax would generally be viewed to be unfair. Most economists would say that such a tax is "horizontally inequitable." More sophisticated economists might distinguish between ex ante horizontal equity and ex post horizontal equity: if the tax were applied in a truly random way, then ex ante it would be horizontally equitable; individuals with identical utility functions and endowments would have identical ex ante expected utility, since they all face equal chances; ex post, it would be horizontally inequitable since individuals with the same endowments and tastes may have very different values of realized

* This paper is a revised version of IMSSS Technical Report #214 (Stiglitz (1976a)). Since this paper was written closely parallel results to those reported in Section 4 were independently derived by L. Weiss (1976). The author is particularly indebted to extended discussions with Richard Arnott and Gary Yohe. I should also like to acknowledge helpful discussions with Frank Hahn, David Newbery, James Mirrlees, Harvey Rosen, David Bevan, Tony Atkinson and Barry Nalebuff. Earlier versions of this paper were presented to the Warwick Summer Workshop (1976), to the mathematical economics seminar at Oxford, and to the public finance seminar at Princeton. I am indebted to the participants at these seminars for their helpful comments. Financial support from the National Science Foundation is gratefully acknowledged.

utility. But most economists would agree with Musgrave (1976) that it is ex post horizontal equity in which we are interested. (These arguments cannot be pushed too far: the draft lottery can be thought of as a random tax applied to a particular subgroup of the population, and it did receive widespread -- though far from universal -- acceptance.)

Indeed, so basic is the notion of horizontal equity that it is incorporated in the Constitution of the United States in the "equal protection clause." The government may not treat differently individuals who are, for the purposes at hand, otherwise identical.

Horizontal equity is usually presented as a principle in its own right. It is not derived from other principles. Nor is there any discussion of the relationship of this principle with other principles. For instance, is it ever inconsistent with Pareto optimality? If it is, does one of the principles have priority over the other?

In recent years there has developed a large literature on optimal tax structures, using a utilitarian (or more general, social welfare) criterion.¹ This approach provides a simple and useful framework within which alternative structures can be evaluated. Most of this literature has, however, ignored the question of horizontal equity.² The question naturally arises, can the principles of horizontal equity be derived from a utilitarian (or more general social welfare) criterion?

¹ See for instance, Ramsey (1927), Diamond and Mirrlees (1971), Stiglitz and Dasgupta (1971), Atkinson and Stiglitz (1972), and Mirrlees (1971). In these papers, the government is assumed to seek, within a certain class of tax structures, that which maximizes, $\sum_j u^j$, where u^j is the utility of the j th individual, or more generally, an individualistic social welfare function of the form $w(u^1, \dots, u^j, \dots)$ where $\partial w / \partial u^j > 0$, and w is a concave function of its arguments. More recently, Stiglitz (1981) has attempted to characterize the set of Pareto efficient tax structures.

² With some exceptions. See Stiglitz (1972), Atkinson and Stiglitz (1976, 1980).

The object of this paper is to show that it cannot.¹ Indeed, we establish that the principle of horizontal equity may be inconsistent with utilitarianism. That is, social welfare (as measured by the sum of utilities) is higher if individuals who have the same tastes and the same endowments are treated differently.² Even more strongly, we show that horizontal equity may be inconsistent with the principle of Pareto optimality.³

Most of our analysis is focused on the desirability of random taxation. We show that random taxation may lead to a Pareto improvement. The implications of our analysis extend, however, to a wide variety of social decisions. Thus, in Section 6 we present several other contexts in which horizontal equity is either inconsistent with social welfare maximization (utilitarianism) or with Pareto optimality; and in Section 9, we discuss briefly the implications of our results for earlier analyses of optimal tax structures.

Our analysis also has implications for pricing policies of monopolists: in Section 8 we show that it may be desirable for regulated and unregulated monopolists to randomize prices.

¹ At least from the form of utilitarianism represented in the optimal tax literature. See below, Section 9.

² The same results may be obtained with other social welfare functions as well.

³ That horizontal equity may not be consistent with Pareto optimality is perhaps not as surprising as it first sounds -- the familiar story of the two shipwrecked sailors with only enough food for one (so horizontal equity entails both dying) at least shows the possibility of a contradiction. We are suggesting that conflict among these principles is more common than such examples may suggest.

We conclude with some speculative remarks about the implications of our results for the role of the concept of horizontal equity in the analysis of questions of public policy.

2. Utilitarianism and Horizontal Equity: The Conventional View

The intuitive derivation of the principle of horizontal equity from utilitarianism is a special application of Lerner's (1944) argument for progressive taxation. Assume two individuals "A" and "B" have identical incomes and utility functions

$$U^A(C^A) \equiv U^B(C^B) \quad \text{if} \quad C^A \equiv C^B$$

where C^i is the i -th individual's consumption:

$$C^i = Y^i - T^i$$

where Y^i is his income and T^i is his tax payment. Thus, if we maximize social welfare

$$\max_{\{T^A, T^B\}} \{U^A(Y^A - T^A) + U^B(Y^B - T^B)\}$$

subject to the revenue constraint

$$T^A + T^B \geq R$$

it is clear that if $U^{1''} < 0$, i.e. there is diminishing marginal utility of income, optimality entails $C^A = C^B$. (See Figure 2.1) Thus, if $Y^A = Y^B$,

$$T^A = T^B,$$

equal taxes paid by identical individuals. Concavity (in this case of the utility function) implies equality and, as most readers of Samuelson's

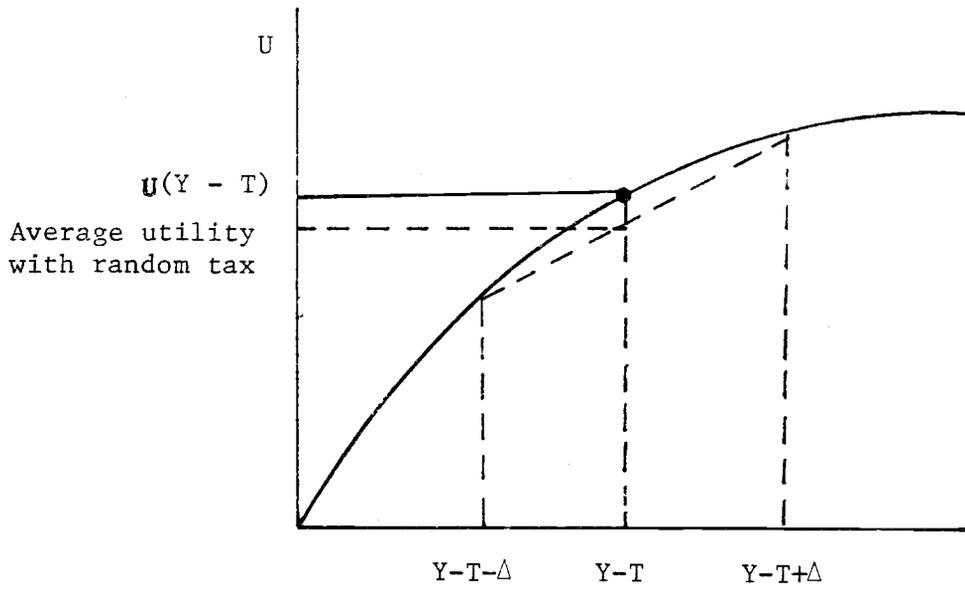


Figure 2.1

With a concave utility function and lump sum taxation, individuals with same Y should face the same taxes.

Foundations will attest, every well behaved problem is concave. The modern corollary of Marshall's dictum that nature abhors discontinuities is that nature abhors non-concavities, and it is this, I suspect, that provides the intuitive rationale for the widespread belief in equality and the belief that the belief in equality can be justified by utilitarianism.¹

It is my belief, on the contrary, that a variety of problems of economic interest exhibit non-concavities of the sort that imply that social welfare maximization may require unequal treatment of equals. Such is the case of indirect taxation to which we now turn.

We consider two versions of the problem of random taxation. In the first, the individual knows his tax rate before he decides on his labor inputs, in the second, he is only told his tax rate after he has supplied his labor (although he knows the probability distribution of tax rates before he decides on his labor supply.) In both cases, randomization may be desirable, although the conditions under which it will be desirable differ in the two cases. In both cases, we focus on the desirability of a small degree of randomization; that is, we provide conditions in which a slight randomization in the tax rate would lead to Pareto improvement (in terms of ex ante expected utility); it should be noted that there may be cases where

¹ Since, as we shall show, this belief is not correct, one can only surmise why someone might have come to such a belief. Talks with economists at a large number of institutions lead me to believe that some argument, such as that given here, though usually slightly less formally presented, lies behind their conclusion.

a slight randomization would not be desirable, but a "large" randomization would. Thus, the case for randomization of taxes is stronger than that presented here.

3. Randomization of Taxes Prior to Labor Decision

3.1 The Model

In this and the next section we consider the simplest possible model of indirect taxation: there is a single good (C) and labor (L). We assume that in the absence of taxation, the wage is unity and the price of output is unity (this is just a normalization) and that output is proportional to labor input. Let τ be the tax rate and p the (after tax) price of consumption goods (relative to labor numeraire). Then¹

$$(3.1) \quad p = 1 + \tau .$$

We write the indirect utility function

$$(3.2) \quad V = V(p, I) = \max U(C, L) \\ \text{s.t. } pC \leq L + I$$

There is no lump sum taxation, and no profits, so income apart from that generated by work is zero;² hence $I = 0$. We can easily derive the individual's consumption function, using Roy's identity,

$$C = C(p, I) = - V_p(p, I) / V_I(p, I)$$

¹ For simplicity, we assume that the price of output remains unchanged throughout the analysis (the production technology is linear.) The results are, however, more general.

² The result extends to the case where there is lump sum taxation but distortionary taxation is also employed. Thus, the results may be extended in a straightforward way to linear income tax schedules. See below, Section 5.

Assume, for simplicity, there are two identical individuals. We wish to maximize social welfare by choosing a probability distribution of tax rates on consumption. We focus on the simplest case where one individual will face a low tax rate and the other individual will face a high tax rate. We randomize the taxes, so each individual has exactly an equal chance of facing the high tax rate $(p^A - 1)$, and the low tax rate $(p^B - 1)$. Thus, his expected utility can be written as

$$(3.3) \quad W = \frac{V(p^A, 0) + V(p^B, 0)}{2}$$

Since the two individuals are identical, maximizing ex ante expected utility is equivalent to maximizing social welfare using any individualistic social welfare criterion. Moreover, if we use a utilitarian criterion, maximizing ex ante expected utility is equivalent to maximizing the sum of (ex post) utilities, i.e.

$$V^A(p^A, 0) + V^B(p^B, 0)$$

where superscripts A and B refer to the different individuals.

In either interpretation, we need to maximize W subject to the constraint that the government raise the requisite revenue:

$$(3.4) \quad (p^A - 1) C^A(p^A, 0) + (p^B - 1) C^B(p^B, 0) \geq R.$$

3.2 Derivation of Sufficient Conditions for Randomization

The indifference curves in (p^A, p^B) space may be concave or convex.

The slope of the indifference curve is

$$(3.5) \quad \left(\frac{dp^A}{dp^B} \right)_{\bar{W}} = - \left(\frac{v_p^B(p^B, 0)}{v_p^A(p^A, 0)} \right)$$

We can calculate the curvature

$$(3.6) \quad \left(\frac{d^2 p^A}{d(p^B)^2} \right)_{\bar{W}} = - \frac{v_{pp}^B(p^B, 0)}{v_p^A(p^A, 0)} + \frac{v_p^B}{(v_p^A)^2} v_{pp}^A(p^A, 0) \left(\frac{dp^A}{dp^B} \right)_{\bar{W}}$$

$$= - \left(\frac{v_{pp}^B(p^B, 0)}{v_p^A(p^A, 0)} + \left(\frac{v_p^B(p^B, 0)}{v_p^A(p^A, 0)} \right)^2 \frac{v_{pp}^A(p^A, 0)}{v_p^A(p^A, 0)} \right)$$

At $p^A = p^B$

$$(3.7) \quad \left(\frac{d^2 p^A}{d(p^B)^2} \right)_{\bar{W}} = - \frac{2}{p} [\rho - \epsilon - 2\eta]$$

where

$\rho \equiv - \frac{v_{II} Y}{v_I}$, the elasticity of marginal utility of income¹

$\epsilon = - \left(\frac{d \ln C}{d \ln p} \right)_{\bar{V}}$, the compensated price elasticity

$Y = pC$, "income"

$\eta = \frac{dC}{dI} \frac{Y}{C}$, "income" elasticity of consumption

The derivation of (3.7) is given in Appendix A.

¹ In risk analysis, this is known as the measure of (relative income) risk aversion. In the analysis of income inequality, it is sometimes referred to as the measure of inequality aversion. See Arrow [1970] Pratt [1964], Atkinson [1970].

The constraint curve (the set of values of p^A and p^B satisfying (3.4)) also may be either convex or concave. Its slope is

$$(3.8) \quad \left(\frac{dp^A}{dp^B} \right)_{\bar{R}} = - \frac{C^B + (p^B - 1)C_{pp}^B}{C^A + (p^A - 1)C_{pp}^A}$$

Using (3.8), we calculate its curvature as

$$(3.9) \quad \left(\frac{d^2 p^A}{dp^{B^2}} \right)_{\bar{R}} = - \left(\frac{2C_{pp}^B + (p^B - 1)C_{ppp}^B}{C^A + (p^A - 1)C_{pp}^A} + \frac{2C_{pp}^A + (p^A - 1)C_{ppp}^A}{C^A + (p^A - 1)C_{pp}^A} \right) \left(\frac{C^B + (p^B - 1)C_{pp}^B}{C^A + (p^A - 1)C_{pp}^A} \right)^2$$

At $p^A = p^B$, we obtain

$$(3.10) \quad \left(\frac{d^2 p^A}{dp^{B^2}} \right)_{\bar{R}} = \frac{2}{p} \frac{2 + \hat{\tau}\nu}{1 - \hat{\tau}(\epsilon + \eta)} (\epsilon + \eta)$$

where

$$\hat{\tau} = \frac{p - 1}{p}, \text{ the percentage tax rate}$$

and

$$v = p \frac{C_{PP}}{C_P}, \text{ the curvature of the demand curve.}$$

At the 45° line, $(dp^A/dp^B)_{\bar{W}} \equiv (dp^A/dp^B)_{\bar{R}} = -1$ so we have either a local maximum or minimum, depending on the relative curvature of the indifference curve and the iso-revenue curve, i.e. on whether

$$(3.11) \quad \rho - \epsilon - 2\eta \gtrless -(\epsilon + \eta) \frac{(2 + \hat{\tau}v)}{1 - \hat{\tau}(\epsilon + \eta)}$$

(See Figure 3.1.b.)

To see what is implied by (3.11), we first observe that if we restrict taxation to "efficient" levels (where increasing tax rates increase revenues) then

$$(3.12) \quad \begin{aligned} \frac{d(p-1)C}{dp} &= C + (p-1) \frac{\partial C}{\partial p} \\ &= C[1 - \hat{\tau}(\epsilon + \eta)] > 0 \end{aligned}$$

Multiplying (3.12) through by $1 - \hat{\tau}(\epsilon + \eta)$ and collecting terms, we establish: a sufficient condition for randomization¹ is that

$$(3.13) \quad \hat{\tau}(\rho - \epsilon - 2\eta - v) > \frac{\rho + \epsilon}{\eta + \epsilon}$$

The larger the revenue to be raised ($\hat{\tau}$) and the more negative the curvature of the demand function, the more likely is random taxation to be desirable.

¹ This is also a necessary condition for the desirability of a "small" randomization; it is possible, however, that although "small" randomizations are undesirable, large randomizations are, as illustrated in Figure 3.1c .

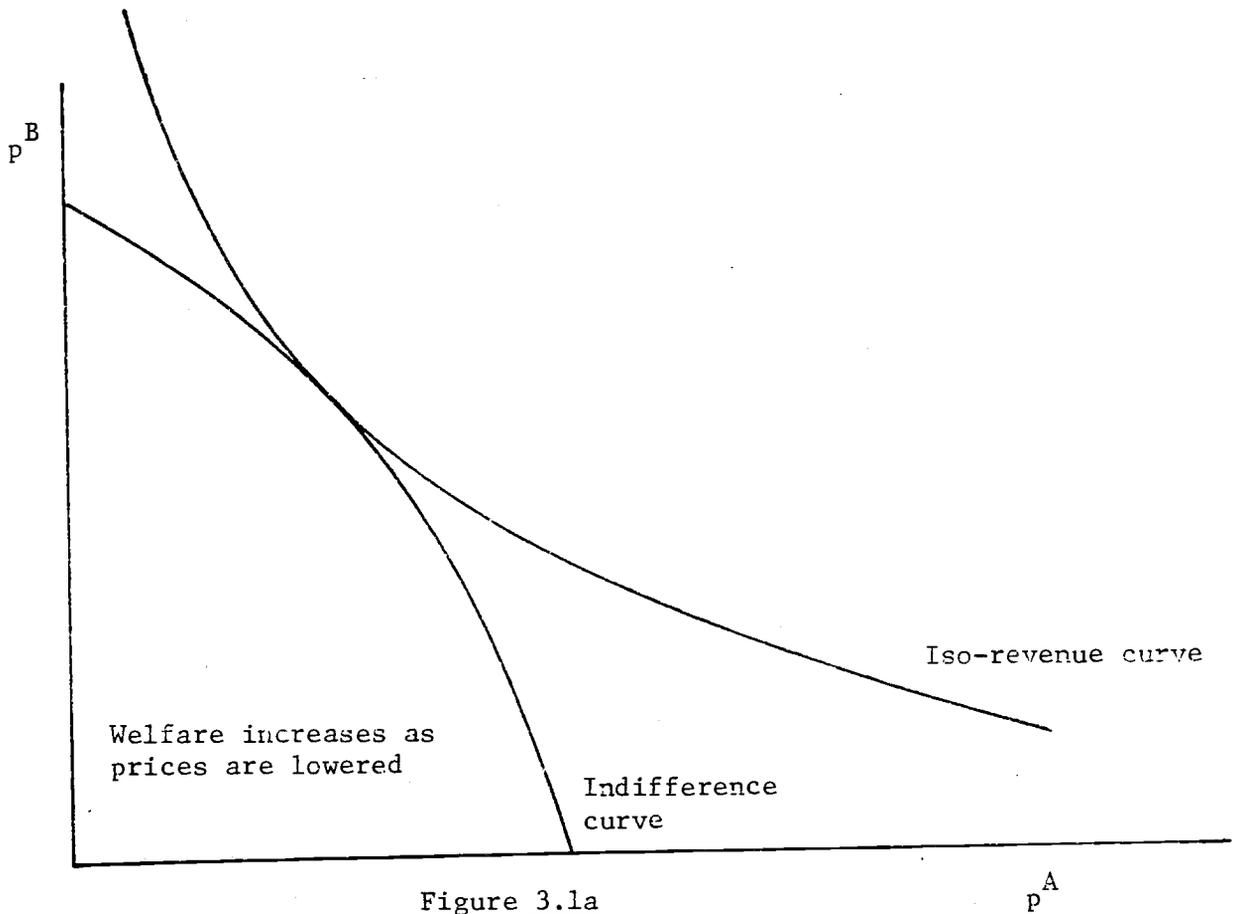


Figure 3.1a

Randomization Not Desirable

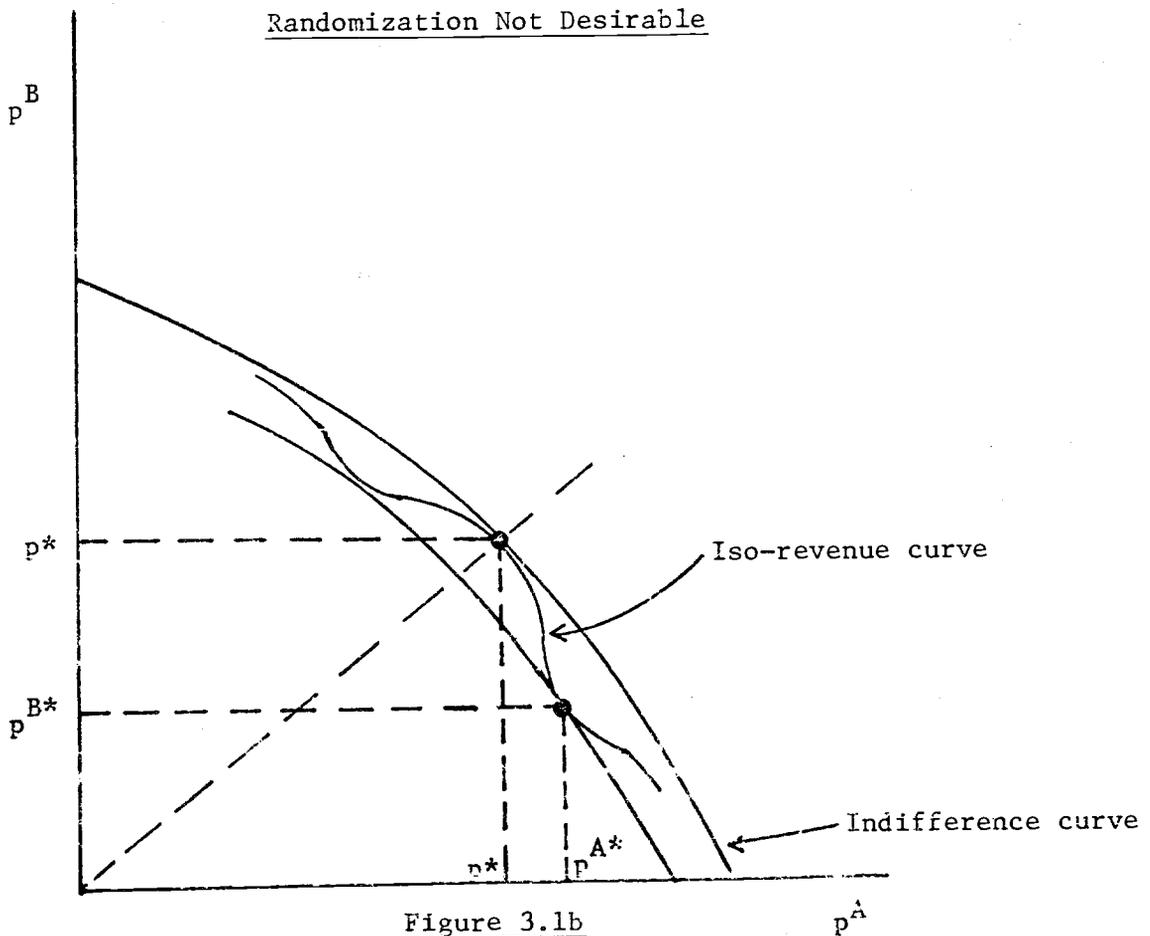


Figure 3.1b

Randomization Is Desirable

Non-randomization (p^* , p^*) is a local minimum. $\{p^{A*}, p^{B*}\}$ is optimal

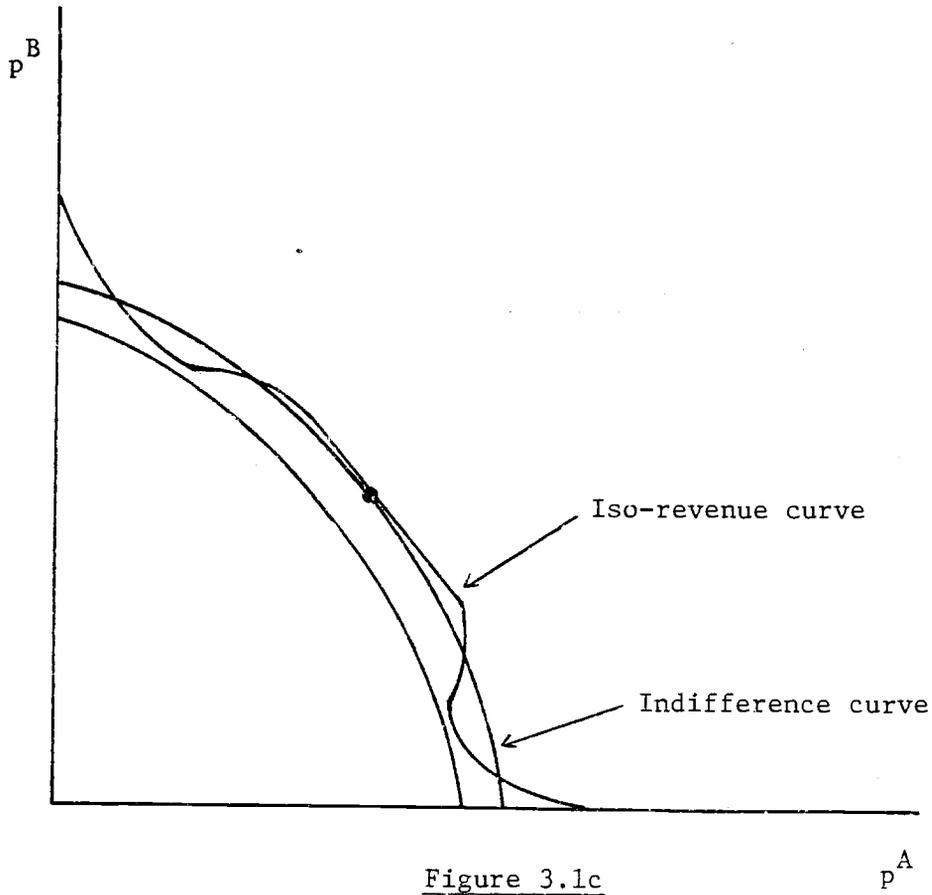


Figure 3.1c

Small Randomization Not Desirable

Large Randomization Desirable

Let us consider the two polar cases:

- (a) At $\hat{\tau} = 0$, randomization is never desirable.
 (b) At the maximum feasible revenue,

$$\hat{\tau} = \frac{1}{\eta + \epsilon} .$$

Hence from (3.12) and recalling the definitions of ϵ , η , and v , we obtain a sufficient condition for the desirability of randomization for sufficiently high values of revenue is¹

$$(3.14) \quad v < -2(\eta + \epsilon)$$

or

$$(3.14') \quad \frac{C_{PP}}{C_P} < \frac{2C_P}{C}$$

3.3 An Example

An example may help illustrate the conditions under which randomization is desirable. Consider the indirect utility function

$$V = \phi \left[-\frac{kp^{1-\beta}}{1-\beta} + \frac{(\bar{L} + I)^{1-\gamma}}{1-\gamma} \right], \quad \phi' > 0$$

which yields the constant elasticity demand functions

$$C = -\frac{V_P}{V_I} = kp^{-\beta} (\bar{L} + I)^\gamma$$

so

$$\epsilon = \beta - \gamma s$$

$$\eta = \gamma s$$

$$\rho = \gamma s + sm > 0$$

$$v = -(1 + \beta)$$

¹ When $\tau = \hat{\tau}$, randomization is, of course, not desirable, since increasing τ lowers revenue.

where

$$s = \frac{pC}{\bar{L} + I}, \text{ the share of "full income" spent on consumption goods.}$$

$$m = - \frac{\phi''}{\phi'} (I + \bar{L}).$$

We require that individuals be risk averse ($m > -\gamma$). Thus, substituting into (3.13) we can show that randomization is desirable provided

$$\frac{sm + \beta}{\beta} < \hat{\tau}(1 + sm)$$

or

$$\hat{\tau} > \frac{1 + sm/\beta}{1 + sm}$$

In particular if the price elasticity of consumption is less than unity

$$\beta < 1,$$

for sufficiently large revenues (so $\hat{\tau} \approx \frac{1}{\eta + \epsilon} = \frac{1}{\beta}$), randomization is desirable.

3.4 Intuitive Interpretation in Terms of Excess Burden

The basic intuition behind our argument can be seen as follows. In Figure 3.2a we have plotted the excess burden (EB) (deadweight loss) imposed on an individual as a function of the revenue raised from him. Clearly, if the curve is concave at the required revenue, \bar{R} , it pays to introduce some randomization, for then average excess burden will be reduced. Thus a sufficient condition for random taxation to be desirable is that

$$\begin{aligned} \frac{d \ln EB}{d \ln R} &= \frac{d \ln EB}{d \ln \tau} \bigg/ \frac{d \ln R}{d \ln \tau} \\ &= \left(\epsilon_{EB, \tau} \bigg/ 1 + \frac{\tau}{p} \frac{d \ln C}{d \ln p} \right) < 1, \end{aligned}$$

where

$$\epsilon_{EB, \tau} = \frac{d \ln EB}{d \ln \tau}, \text{ the elasticity of excess burden with respect to the tax rate.}$$

This can be rewritten equivalently as

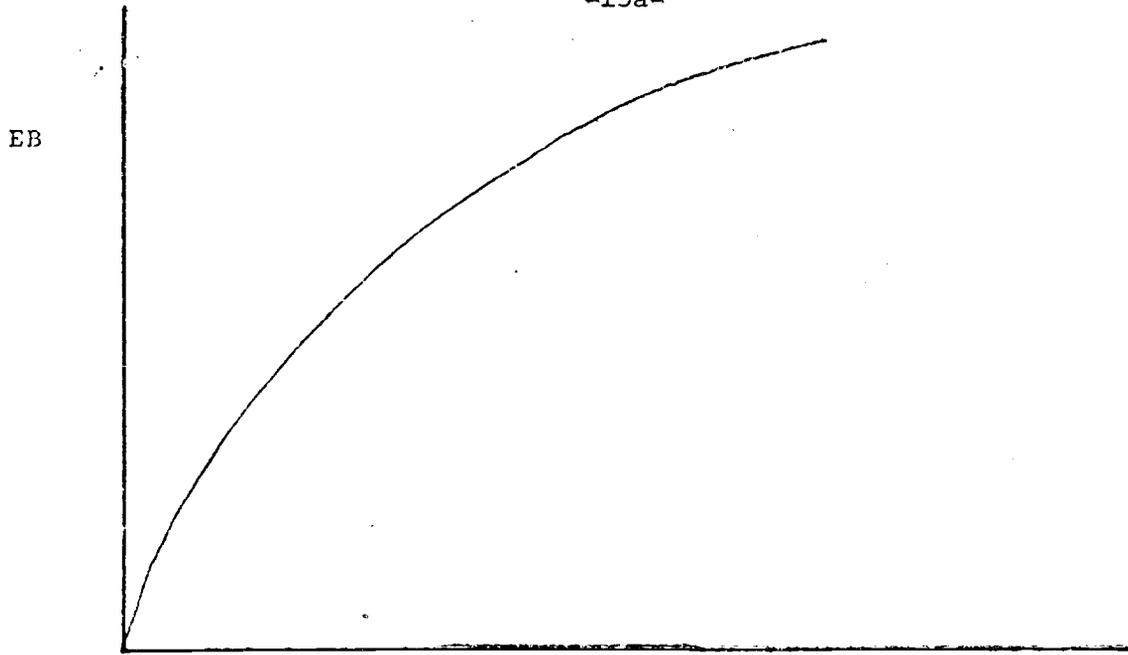
$$\epsilon_{EB, \tau}^{-1} < \frac{\tau}{p} \frac{d \ln C}{d \ln p}$$

As Figure 3.2b illustrates, the percentage increment in excess burden from an increase in the tax rate will be small if the consumption demand curve is convex.

These conditions can perhaps be interpreted more easily in terms of a tax on labor. Let t be the tax on labor, and L be labor supply. Then, we obtain as before that randomization is desirable if

$$\frac{d \ln EB}{d \ln R} = \frac{d \ln EB}{d \ln t} \bigg/ \frac{d \ln R}{d \ln t} < 1$$

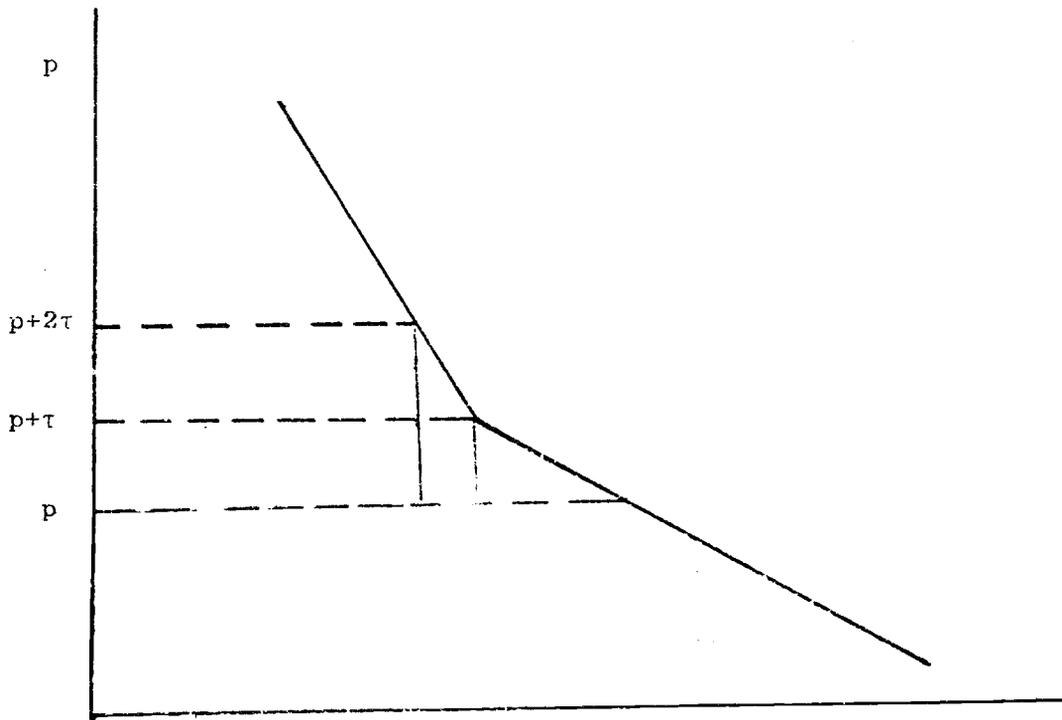
Thus some randomization is desirable if



R

Figure 3.2a

Excess Burden as Function of
Revenue Raised



C

Figure 3.2b

Relationship Between Marginal Excess Burden and Marginal
Tax Depends on Shape of Demand Curve

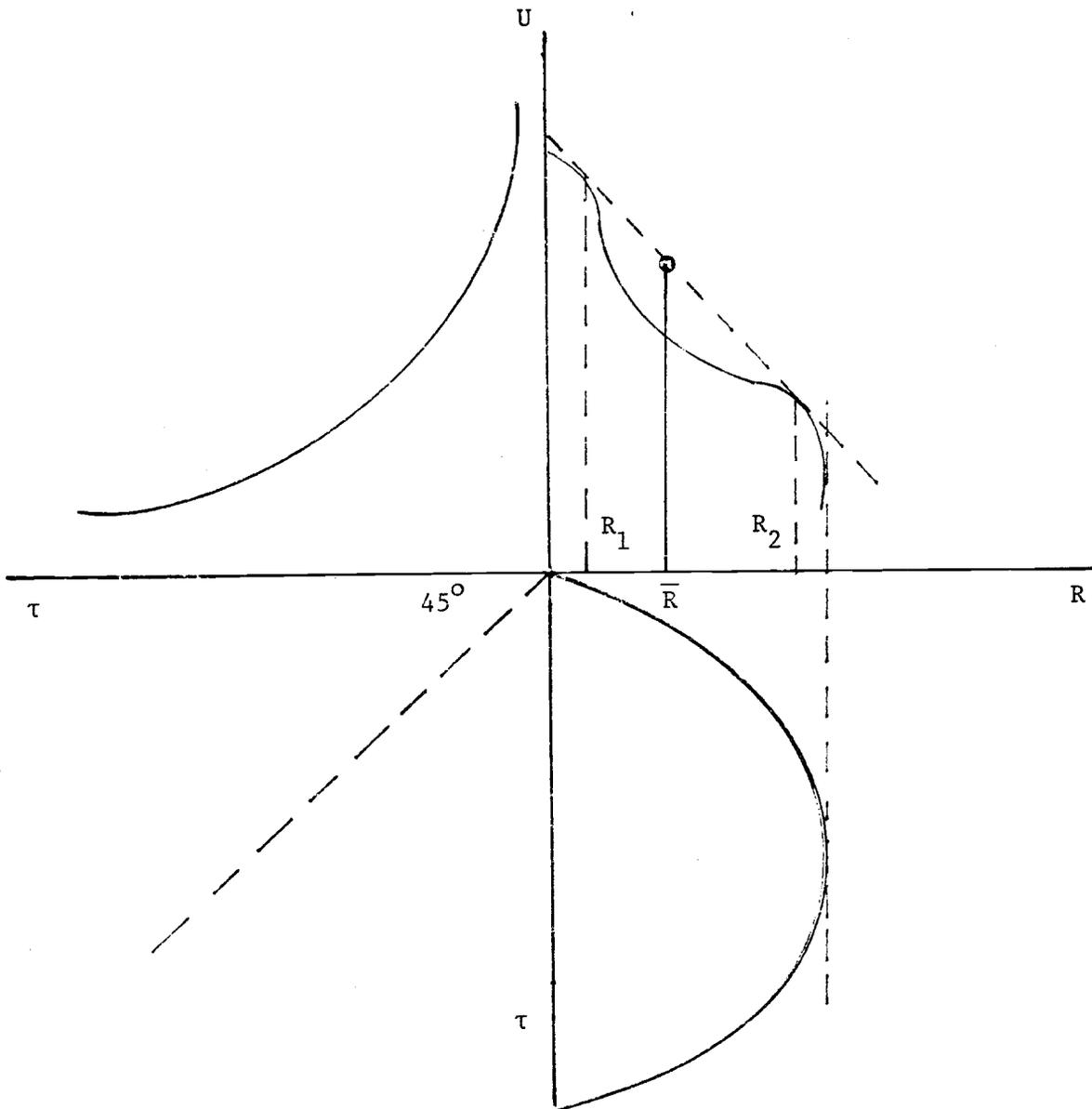


Figure 3.3.

$$\frac{d \ln EB}{d \ln t} - 1 < \frac{d \ln L}{d \ln t}$$

the elasticity of the excess burden with respect to the tax rate is less than one plus the elasticity of the labor supply with respect to the tax rate.

Notice that it is the uncompensated elasticities which are relevant here, for it is the uncompensated elasticities which are critical in determining the shape of the revenue function.

3.5 An Alternative Interpretation

There is an alternative interpretation that will prove useful in some of the subsequent discussion. In Figure 3.3 we have depicted the relationship between the revenue raised from an individual (by means of a proportional consumption tax) and the utility he attains. (The curve is derived from plotting, in the lower right hand quadrant, the relationship between the revenue raised and the tax rate, and in upper left hand quadrant, the relationship between the utility attained and the tax rate.) This utility-revenue curve may not be concave; clearly, if it is not, we can increase average utility by concavifying the curve (as in the diagram). To collect the average revenue \bar{R} , it is optimal to collect the revenue R_1 from some individuals and the revenue R_2 from others.

3.6 The Optimal Randomization Scheme

So far, we have established that some randomization is preferable to no randomization. We now analyze the optimal random tax structure. Let $F(\tau)$ be the proportion of individuals

assigned a tax rate less than or equal to τ . Then we seek that $F(\tau)$ function which

$$(3.15) \quad \text{maximizes } \int V(1 + \tau, 0) dF(\tau)$$

s.t.

$$(3.16a) \quad \int \tau C(1 + \tau, 0) dF(\tau) \geq \bar{R}$$

$$(3.16b) \quad \int dF(\tau) = 1.$$

Letting μ and γ be the Lagrange multiplier associated with the constraints, we obtain

$$(3.17) \quad V(1 + \tau, 0) + \mu \tau C(1 + \tau, 0) = \gamma \quad \text{for all } \tau \text{ with positive density}$$

$$V(1 + \tau, 0) + \mu \tau C(1 + \tau, 0) < \gamma \quad \text{otherwise.}$$

We now prove there exists an optimal probability distribution of positive density at at most two points.

Assume not. There are then at least three tax rates, $\tau_1 < \tau_2 < \tau_3$ with relative frequency π_i , $\sum \pi_i = 1$, yielding revenues (per capita) of R_i , for an average revenue \bar{R} , with $R_1 < \bar{R} < R_3$, and yielding utility levels V_i , with average utility level \bar{V} , $V_1 > \bar{V} > V_3$. From (3.17), V_i is a linear function of R_i . Hence the same level of expected utility could be attained simply by randomizing among τ_1 and τ_3 , with

$$\pi_1 = \frac{R_3 - \bar{R}}{R_3 - R_1}.$$

The result is obvious, of course, from the concavification of the utility-revenue curve illustrated in Figure 3.3.

3.7 Randomization of Optimal Linear Income Taxes

In the simple model we have developed here, with identical individuals, there is no real reason to impose a distortionary tax: a uniform lump sum tax would clearly be preferable. It is only because individuals differ, say, in their abilities, but these differences are not directly observable, that we need to resort to distortionary taxation. (See Atkinson and Stiglitz (1976, 1980).) Our analysis can be easily extended to show that the optimal linear tax structure involves randomization of the marginal tax rate on consumption.

To see this, assume we have a distribution of individuals by ability (before tax real wage) given by $G(w)$.¹ With a linear tax structure, the individual's budget constraint is given by

$$C = I + (1 - t)wL$$

where I is the lump sum payment to each individual and t is the marginal tax rate. The individual's utility is represented by his indirect utility function, now written as a function of the after tax wage rate and the lump sum payment,

$$\hat{V} = \hat{V}(w(1 - t), I)$$

Assume the government can impose different marginal tax rates randomly; as before, half the population faces a rate of t^A , half a rate of t^B . The government wishes to

¹ For a more extended discussion of optimal linear tax structure, in the absence of randomization, see Stiglitz (1976c).

$$\text{maximize}_{\{I, t^A, t^B\}} \frac{1}{2} [\hat{V}(w(1 - t^A), I) dG(w) + \hat{V}(w(1 - t^B), I) dG(w)]$$

subject to the budget constraint

$$\frac{1}{2} \int [t^A w L(w(1 - t^A), I) + t^B w L(w(1 - t^B), I)] dG(w) = I + \bar{R}$$

where \bar{R} is the government's expenditure (per capita) on public goods (taken to be fixed).

The analysis proceeds exactly as before. We take (for the moment) I to be fixed. Then $t^A = t^B$ is always a critical point, but it may be a local minimum rather than a local maximum. We can derive expressions analogous to (3.13) and (3.14) -- but now involving appropriately weighted averages of the demand elasticities, risk aversions, etc. -- providing sufficient conditions for randomization.

There is one problem with the implementation of the kind of random tax scheme we have described in this section.

Since whether the individual will be faced with a low or a high tax rate is an insurable risk with no moral hazard associated with it, clearly individuals would be willing to purchase insurance to reduce this risk. If perfect insurance were purchased, the individual's behavior would be identical to that with no randomization, and obviously then randomization would have no effect. Thus, it is apparent that a social

optimality requires, in this case, restrictions on the set of insurance markets which are allowed to operate.¹

4. Randomization of Taxes After Labor Decision

The reason that random taxation was desirable in the previous section was, roughly, that the amount of revenue raised increased more than proportionately to the tax because of the differences in response of labor supply to different after tax wages. Thus, the average tax rate could be reduced by having some individuals face a high tax rate and some a low rate.

In this section, we consider the case where the individual must decide on his labor input prior to knowing the tax. If the individual is risk averse, he will "plan" on facing a high tax rate, and hence each individual will reduce his labor supply by less than he otherwise would; this enables the average tax rate to be reduced. Individuals are

¹ There may be no scope for insurance of the conventional kind (the individual pays so much to the insurance company if his tax rate is low, receiving some fixed amount if his tax rate is high). Whether such policies are desirable, and the nature of these policies, depends on the value of V_{pI} . If $V_{pI} = 0$, as it may (the marginal utility of income does not depend on the real wage), then there is no scope for such insurance contracts. If $V_{pI} < 0$, then the insurance contract actually leads the individual with the higher tax rate to pay money to the individual with the lower tax rate. With normal demand curves, this would decrease the desirability of randomization of a consumption tax, since it will redistribute consumption away from highly taxed individuals. The insurance contracts may make randomization undesirable, and the government may need to intervene to restrict such insurance markets.

worse off because they face the risk arising from the random tax. They are better off because they face a lower average tax rate. We shall show that this second effect can dominate the first; randomization may increase everyone's ex ante expected utility.

It is more convenient in this section to take the tax as one on labor; the individual is assumed to face the tax rate $t + \Delta$ with probability .5, and $t - \Delta$ with probability .5, where $\Delta \geq 0$. Now, we take the price of output as our numeraire ($p = 1$). The individual chooses L to

$$(4.1) \quad \text{maximize} \quad \frac{U((\omega + \Delta - t)L, L) + U((\omega - \Delta - t)L, L)}{2} \equiv W$$

where ω is the real wage rate. The revenue constraint is now

$$(4.2) \quad tL = \bar{R}.$$

Since individuals are all identical, maximizing the individual's expected utility is equivalent (as before) to maximizing social welfare.

In Appendix B, we show that

$$(4.3) \quad \left(\frac{dW}{d\Delta} \right)_{\Delta=0} = 0$$

but that (differentiating (4.2) again)

$$(4.4) \quad \left(\frac{d^2W}{d\Delta^2} \right)_{\Delta=0} = U_{11}L^2 - U_1L \frac{d^2t}{d\Delta^2}$$

$$= - \frac{U_1L}{\omega-t} \left[\rho + \frac{d^2t}{d\Delta^2} (\omega - t) \right]$$

where, **now** we define

$$(4.5) \quad \rho = \frac{-U_{11}C}{U_1}$$

Hence, as Figure 4.1 illustrates randomization is desirable if

$$(4.6) \quad \left. \frac{d^2 t}{d\Delta^2} \right|_{\bar{R}=0} = \frac{-\frac{\partial^2 L}{\partial \Delta^2} \frac{t}{L}}{1 + \frac{\partial L}{\partial t} \frac{t}{L}} < \frac{-\rho}{\omega - t}$$

The reduction in the tax rate depends on three factors:

(a) The responsiveness of labor supply to risk. Variability in after tax wages increases or decreases labor supply (Rothschild-Stiglitz (1971)), depending on the concavity or convexity of the first order condition ($U_1\omega + U_2 = 0$) in terms of the wage ω . (The change in the after tax wage is a mean-preserving spread in the wage distribution.) Clearly, a necessary condition for the desirability of randomization is that risk increases labor supply ($\partial^2 L / \partial \Delta^2 > 0$). (Later we provide conditions ensuring that this will occur.)

(b) The responsiveness of labor supply to taxes. The more that an increase in taxes reduces labor supply, the greater the return from being able to reduce the (average) tax rate, and thus the more likely that randomization will be desirable.

(c) The size of the required tax revenues \bar{R} . When $t = 0$, $\frac{d^2 L}{d\Delta^2} = 0$; hence for sufficiently small tax revenues randomization is never desirable.

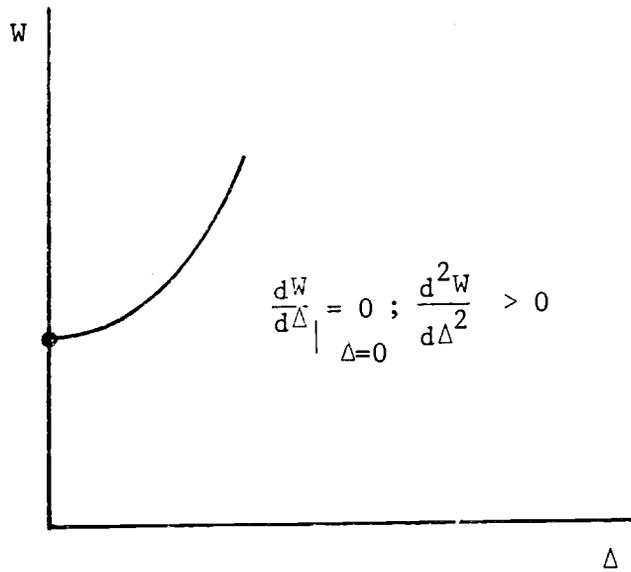


Figure 4.1

Randomization is Desirable

if $\frac{d^2W}{d\Delta^2} > 0$.

To see the effect of large revenues, we observe that the denominator of (4.6) can be rewritten as

$$(4.7) \quad 1 + \frac{\partial L}{\partial t} \frac{t}{L} = \frac{1}{L} \frac{dtL}{dt} = \frac{1}{L} \frac{d\bar{R}}{dt}$$

The value of t which maximizes revenues without randomization is denoted by t^* , and the corresponding value of \bar{R} by R^* . Thus,

$$(4.8) \quad \hat{t}^* \equiv \frac{t^*}{w} = 1 / \frac{\partial \ell_{nL}}{\partial \ell_{nw}} = 1/\text{elasticity of labor supply.}$$

The maximal tax rate is the inverse of labor supply elasticity.

Thus, if risk increases the labor supply and if there exists a maximal revenue without randomization of R^* , for sufficiently high revenues, randomization is desirable, since

$$\left. \frac{d^2 t}{d\Delta^2} \right|_{\Delta=0} \rightarrow -\infty \quad \text{as } t \rightarrow t^*.$$

Indeed, for sufficiently large government expenditures, the only way of raising the requisite revenue may be to randomize. For, if at t^* , randomization increases labor supply, clearly government revenue will be raised.

4.1 Derivation of Sufficient Conditions for Randomization in Terms of Utility Functions

To see more generally the conditions under which randomization is desirable, we need to express $d^2 t/d\Delta^2$ in terms of the utility function. Straightforward differentiating of the first order conditions yields

$$(4.9) \quad - \frac{\partial^2 L}{\partial \Delta^2} \Big|_{\Delta=0} = \frac{U_{111} L^2 (\omega - t) + 2U_{11} L + U_{211} L^2}{U_{11} (\omega - t)^2 + 2U_{21} (\omega - t) + U_{22}}$$

and

$$(4.10) \quad \left. \frac{\partial L}{\partial t} \right|_{\Delta=0} = \frac{U_{11}L(\omega - t) + U_{12}L + U_1}{U_{11}(\omega - t)^2 + 2U_{21}(\omega - t) + U_{22}}$$

Substituting (4.9) and (4.10) into (4.6) and that into (4.4), we obtain

$$(4.11) \quad \left. \frac{d^2 W}{d\Delta^2} \right|_{\Delta=0} = \frac{-U_1 L}{1 + \frac{\partial \ln L}{\partial \ln t}} \frac{t U_{111} L(\omega - t) + 2t U_{11} + t L U_{211}}{U_{11}(\omega - t)^2 + 2U_{21}(\omega - t) + U_{22}} + U_{11} L^2$$

The denominator of (4.9)-(4.11) is unambiguously negative (provided that the tax rate is below that which maximizes revenue). Hence a sufficient condition for the desirability of randomization is that (from (4.4)), $\frac{d^2 t}{d\Delta^2}$ be large (and negative) relative to ρ ; from (4.6) this will be the case if $\frac{\partial^2 L}{\partial \Delta^2}$ is large (and positive), which from (4.9), will be true if U_{111} is positive and large. It is clear not only that U_{111} can be positive, but it can be very large, in which case randomization will be desirable (for t large enough).

4.2 Separable Utility Functions

To get a better idea of the kinds of conditions under which randomization might be desirable, assume we had a separable utility function, so $U_{21} = U_{211} = 0$. We define

$$(4.12) \quad \xi = - \frac{U_{22}L}{U_2}$$

and

$$(4.13) \quad -\Lambda \equiv \frac{C\rho'}{\rho} = \frac{U_{111}C}{U_{11}} + 1 - \frac{U_{11}C}{U_1}$$

$$= \frac{U_{111}C}{U_{11}} + (1 + \rho)$$

Λ is the elasticity of ρ , the individual's risk aversion.

Substituting into (4.9) and (4.10) we obtain

$$(4.14) \quad - \frac{\partial^2 L}{\partial \Delta^2} \Big|_{\Delta=0} = \frac{\rho}{\rho + \xi} \frac{L}{(\omega - t)^2} (-\Lambda + 1 - \rho)$$

$$(4.15) \quad \frac{\partial L}{\partial t} = \frac{(\rho - 1)L}{(\omega - t)(\rho + \xi)}$$

Thus, randomization leads to increased labor supply if

$$\Lambda > 1 - \rho$$

and increased taxation reduces the labor supply if

$$\rho < 1.$$

Substituting (4.14) and (4.15) into (4.6) we observe that

$$(4.16) \quad \frac{d^2 W}{d\Delta^2} \Big|_{\Delta=0} \geq 0 \text{ as}$$

$$\hat{t} \equiv \frac{t}{\omega} \geq \frac{\xi + \rho}{\Lambda + \xi + \rho}$$

Thus, with separability, if there is rapidly decreasing relative risk aversion (Λ is very large), randomization becomes desirable even at low tax rates.

These results, like those of the preceding section, may seem rather counterintuitive: after all, one is imposing more risk on the individual. Yet, remarkably enough, the condition we have derived (in the simple case of separable utility functions) suggests that randomization may be attractive even with high risk aversion.

There are two effects of randomization. First, it imposes a risk on individuals. By the usual kinds of arguments, the welfare loss can be shown to be of the order of (for small risks) $\rho\Delta^2/2$. At the same time, it affects labor supply; the change in aggregate labor supply affects the (average) tax rate; changing the average tax rate changes the deadweight loss associated with the tax. The deadweight loss is approximately

$$\frac{t^2}{2} \frac{\partial L}{\partial \omega} = \frac{\bar{R}^2}{2L^2} \frac{\partial L}{\partial \omega} = \frac{1}{\omega L - \bar{R}} \cdot \frac{\bar{R}^2}{2} \left(\frac{\omega-t}{L} \frac{\partial L}{\partial \omega} \right)$$

Thus, the deadweight loss is inversely related (for constant elasticity supply functions of labor) to the aggregate labor supply. If aggregate supply increases when we randomize, i.e. if $\frac{\partial^2 L}{\partial \Delta^2} > 0$ is sufficiently large, then the effect of the benefit from reduction in the average tax rate is greater than the loss from the induced risk.

One interpretation of the kind of random taxation we have discussed in this section is the random enforcement of taxes. (Fixed costs of auditing increase the desirability of random audits.) This interpretation has been discussed at greater length by Weiss (1976) .

5. The Randomization of the Optimal Tax and Optimal Randomization

In the preceding two sections we established the desirability of randomization for linear tax structures. In this section, we ask, what can we say about optimal randomization of linear tax structures, on the one hand, and the randomization of optimal non-linear tax structures, on the other.

5.1 Optimal Ex Post Randomization

In Section 3.6 , we established that the optimal ex ante randomization (i.e. randomization before the individual has decided on his level of effort) required only two tax rates.¹ Here, we show that ex post randomization entails only three tax rates.

Formally, the government wishes to find a probability distribution of tax rate represented by $F(t)$, which maximizes individuals' expected utility subject to the government's budget constraint, i.e.

¹ The generalization of that result to n commodities and labor requires randomization among $n + 1$ tax structures.

$$(5.1) \quad \max_L [\max \int U[(\omega - t)L, L] dF(t)]$$

$$(5.2a) \quad \text{s.t.} \quad L \int t dF(t) \geq \bar{R}$$

$$(5.2b) \quad \int dF(t) \equiv 1.$$

A necessary condition for this is that we maximize expected utility, given the labor supplied, i.e. from the individual's first order conditions for optimal L , we have

$$(5.3) \quad \int [U_1(\omega - t) + U_2] dF(t) = 0.$$

Letting μ be the Lagrange multiplier associated with the revenue constraint (5.2a), γ be the Lagrange multiplier associated with the constraint (5.2b), and ψ be the Lagrange multiplier associated with the constraint (5.3), then

$$(5.4a) \quad U((\omega - t)L, L) + \mu L + \psi [U_1(\omega - t) + U_2] = \gamma \quad \text{if } t \text{ occurs with positive probability}$$

$$(5.4b) \quad U((\omega - t)L, L) + \mu L + \psi [U_1(\omega - t) + U_2] < \gamma \quad \text{otherwise}$$

To see that there need be at most three tax rates which occur with positive probability, assume the contrary, i.e. $t_1 > t_2 > t_3 > t_4$ all occur with positive probability. Letting π_i = the relative frequency of t_i , $\sum_i \pi_i = 1$.

Define

$$(5.5) \quad \phi(t) \equiv U_1(\omega - t) + U_2$$

$$(5.6a) \quad \bar{\phi} = \sum_i \pi_i \phi(t_i)$$

$$(5.6b) \quad \bar{U} = \sum_i \pi_i U((\omega - t_i)L, L)$$

If $\{t_1, t_2, t_3, t_4\}$ do in fact occur with relative frequency π_i in the optimal random tax, it must be the case that $\{\pi_i\}$ is the solution to

$$(5.7) \quad \max_{\{\pi_i\}} \sum t_i \pi_i$$

$$(5.8a) \quad \text{s.t.} \quad \sum \phi(t_i) \pi_i = \bar{\phi}$$

$$(5.8b) \quad \sum U((\omega - t_i)L, L) \pi_i = \bar{U}$$

Thus reformulated, this is simply a linear maximization problem subject to two linear constraints, and the result is immediate.

5.2 Ex Post and Ex Ante Randomization

In the two preceding sections, we analyzed separately ex post and ex ante randomization. In fact, the optimal tax structure may entail both simultaneously; that is, individuals are told, before they decide on the level of effort, that they will face one of two random tax lotteries, one yielding, say, $t^A - \Delta^A$ with probability p^A and $t^A + \Delta^A$ with probability $1 - p^A$, the other yielding $t^B - \Delta^B$ with probability p^B and $t^B + \Delta^B$ with probability $1 - p^B$. Given that ex post randomization is desirable, the desirability of ex ante randomization may be analyzed exactly as before, but now, for each average level of tax, we calculate the optimal random distribution and the associate level of expected utility and average revenue.

5.3 Randomization with Optimal Taxes

This paper has focused on the desirability of randomization when the government is restricted to employing linear taxes. The question naturally arises, is randomization desirable if this restriction is removed? Does randomization, for instance, arise simply because of the second best nature of the problem?

On the contrary, it turns out that Pareto efficient taxation, with individuals differing say by ability as in the model of Section 3.7, entails randomization under much weaker conditions than those derived in the preceding two sections. The use of randomization enables the government to distinguish between high and low ability individuals, at a lower cost (in terms of the distortions imposed). The analysis of this problem involves rather different techniques than those employed here, and hence is taken up elsewhere. (See Stiglitz (1981).)¹

6. Other Contexts Where Utilitarianism Implies Horizontal Inequity

There have been several other contexts in which utilitarianism implies horizontal inequity:

(a) The efficiency wage hypothesis: Mirrlees (1975) and Stiglitz (1976b) have analyzed the optimal distribution of income in a family farm in which the productivity of the individual depends on the wage he receives (the amount of food he consumes), as in Figure 6.1.

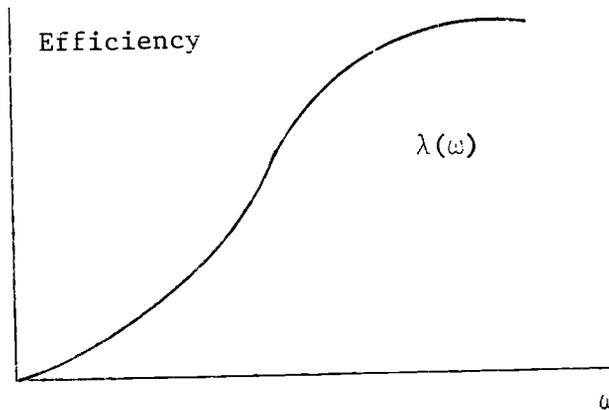


Figure 6.1

¹ This result is, in fact, a special case of a more general theorem about the desirability of randomization in principle agent problems. See Stiglitz (1982).

$\lambda(\omega)$ gives the number of efficiency units supplied by an individual receiving a wage ω . Optimality in general entails some individuals receiving a low wage, some a high wage. Indeed, introducing inequality may be ex post Pareto optimal. When everyone receives the same wage, they are all unproductive, and output is low. Giving a few individuals a lot more enables them to work so hard that not only do they produce enough to provide the extra food that they consume, but they have some left over to give to the remaining low wage individuals.

(b) The optimal town: Mirrlees (1972) has shown that in the optimal town the utilitarian solution entails inequality: individuals who are randomly assigned to live further from the center enjoy more land and have a higher level of utility. By crowding individuals near the center, the transport costs of all those further out are reduced, and this gain in efficiency more than offsets the inequality generated thereby.

(c) Fixed training costs:¹ Assume everyone is ex ante the same, and there are two types of jobs to be performed. One requires training costs T (denominated in say the labor numeraire). Thus, some individuals are (randomly) assigned to the high training cost industry, others to the low training cost industry. It can be shown that those individuals who are (randomly) assigned to the industry requiring training will, in the utilitarian optimum, work harder and have a lower level of utility. By making these individuals work harder, expenditures on training costs are reduced; in effect, net national product can be increased, at the cost of some increase in inequality. Unless society has infinite inequality aversion (using the Atkinson measure of inequality), it always pays to introduce some inequality.

¹ This problem was discussed in Stiglitz (1973). A formal derivation of this result is contained in Stiglitz (1976a).

These examples have one feature in common: they lack, in one way or another, the concavity property which is required for utilitarianism to imply horizontal equity. In the first ("efficiency wage") problem, inequality in income raises net national product by increasing average productivity; in the third (training cost) problem, inequality in the labor required of different individuals raises net national product by reducing the expenditure on training cost; in the second (optimal city) problem, inequality in the allocation of land raises net national product by reducing total expenditure on transport costs.

In some situations, lifetime equality may be attained, even though there is, at any moment, inequality. Thus, in the optimal town problem, if it were costless to move individuals, we could rotate individuals among plots of land so lifetime utilities were identical. But this is not possible for the training cost problem. In the optimal tax problem, if individuals can save, then again randomization may entail lifetime inequality.

In all of the situations examined, randomization is Pareto optimal: ex ante expected utility is maximized by this kind of inequality.

7. Markets and Inequality

It is thus not surprising that in situations such as those discussed in this paper but arising in market contexts, there will be a tendency for the market to introduce randomization. For instance, consider the problem of financing specific training costs. The efficient way for the firm to recover those training costs is to impose, in effect, a lump sum tax on all workers, but this may not be desirable; if individuals do not

know their ability prior to training, they would prefer an "income tax," i.e. a piece rate less than the value added.¹ Assume for administrative reasons that the piece rate cannot be made a function of the number of items produced (i.e. we must have a linear income tax.)² Then equilibrium may entail a random piece rate. Perhaps this provides part of the explanation of random promotion policies (equivalent to random wages) used by some universities.

Note that so long as individuals have a free choice of occupations, the wage contract must entail individuals who are ex ante identical receiving the same level of expected utility.

(Similarly, a developer developing a new residential town and selling off the plots of land, could be better off (i.e. increase his profits) by having a fixed fee for a plot, having the plots of different sizes, and randomly assigning individuals to a plot.)

In the kind of situations we have depicted, the market allocation may not be Pareto efficient for two reasons. First, it may be difficult to enforce random contracts; in the example described earlier, with fixed training costs, the worker who receives training works harder, but receives the same wage, as the one who does not receive training. In such a situation, there is obviously an incentive for some other firm to bid the worker away. If it is not possible to restrict labor mobility, then the market equilibrium

¹ The parallel between "income taxes" and "piece rates" is discussed at greater length in Stiglitz (1975).

² This assumption is not essential, it is made only to convert the problem at hand directly into one which is equivalent to that analyzed in Section 3. More generally, it would appear that even if non-linear piece rates were admissible, randomization may be desirable, but we have not analyzed the conditions under which this will be true.

will require that all (ex ante identical) individuals receive the same level of utility (ex post); the market, in these situations, entails excessive egalitarianism.

Second, in the training cost problem presently under consideration, the optimal resource allocation requires, in effect, subsidies from the trained workers to the untrained, or conversely. As we noted earlier, if there were a single government-controlled firm, it would pay equal wages to all workers, but randomly assign some workers to the plant not requiring training and some to the plant requiring training. It would require that the latter workers work harder (longer). Although the workers so assigned might well complain, the ex ante expected utility of all workers is maximized by this randomization. But the utilitarian allocation is not, in general, viable under competition. For in general, the revenues raised from the sale of one commodity will not be equal to the expenditures on wages plus training costs: there will be a subsidy in one direction or the other. It is obvious, then, that the equilibrium cannot be sustained by a laissez-faire competitive equilibrium.¹

8. Random Pricing by Regulated and Non-Regulated Monopolies

It is by now well recognized that there is a close relationship between the analysis of optimal tax structures (Ramsey and his descendants) and the analysis of optimal pricing of public utilities (Boiteux and his descendants). Thus, our remarks about the desirability of random taxes apply directly to the problem of pricing of regulated utilities.

¹ For proof, see Stiglitz (1976a).

But our analysis also shows that there are conditions under which an unregulated monopolist may find randomization desirable. We noted earlier, for instance, that the maximization of government revenue may entail randomization of prices (tax rates). Thus, if production were controlled by a single monopolist, he would, in these circumstances, randomize his prices.

Similarly, the optimal entry deterring strategy for a monopolist may entail randomizing its prices.¹

Consider a monopolistic firm wishing to maximize its revenue; assume it has a long term contract with its customers; the contract specifies the "price distribution" which it will charge (the commodity is assumed not to be storable). The customer has a reservation expected utility level, i.e. at any contract yielding a lower level of expected utility, he purchases a substitute for the given commodity.²

Then the problem of the monopolist is maximizing revenue subject to this expected utility constraint as opposed to our problem, maximizing expected utility, subject to a revenue constraint. Formally the two problems are simply dual to each other. Thus under perfectly analogous conditions to those presented here, the monopolist will find it desirable to randomize his prices.

¹ See Salop (1979), Dasgupta and Stiglitz (1980) and Gilbert and Stiglitz (1979) for more general discussions of entry deterrance.

² Newbery (1978) has explored one version of this application of our general model.

9. Implications of our Analysis for the Theory of Optimal Taxation
and the Usefulness of the Concept of Horizontal Equity

Most of the literature in the theory of optimal taxation (surveyed in Atkinson and Stiglitz (1980)) has focused on characterizing first order conditions for the maximization of social welfare. Our analysis has made it clear that this may be insufficient: under not unreasonable circumstances these may characterize a local minimum rather than a maximum.

Although we have formulated the problem in terms of two absolutely identical individuals, the issues we have raised arose, in slightly disguised form, in the standard **optimal** tax problem. Consider, for instance, two individuals one of whom likes brown ice cream (but obtains zero utility from white), the other of whom likes white ice cream. Assume they have identical demand functions. Assume moreover that the costs of production of white and brown ice creams are identical. Thus, if τ_b is the tax rate on brown ice cream, τ_w that on white ice cream, then in the conventional optimal tax problem, there may be three critical points, at $\tau_b = \tau_w$, at $\tau_b > \tau_w$, and at $\tau_b < \tau_w$. One might well consider solutions with $\tau_b \neq \tau_w$ horizontally inequitable, but such a solution may be a local utilitarian maximum. The optimal tax structure is, in any reasonable sense, horizontally inequitable.

Many economists would, accordingly, reject the asymmetric solution in which the two colors of ice cream are treated differently. Presumably, a true believer in utilitarianism would not.

But let us now assume that there are, within the population, two groups of individuals distinguished, say, only by hair color. Recall that we showed earlier then there were two alternative interpretations

of our results. We focused on the situation where we treated all individuals ex ante identically, and thus randomization was (ex ante) Pareto optimal; but even if type A individuals always were taxed at a higher rate, and type B individuals always were taxed at a lower rate, social welfare (using the utilitarian criterion) was increased as a result of taxing them at different rates. A strict utilitarian would thus have no way of choosing between a tax system which randomized taxes (resulting in identical ex ante expected utility) and one which always taxed light haired individuals at a lower rate than dark haired individuals; moreover, under the conditions provided in the earlier analysis, he would argue that a tax system which differentiated between light haired and dark haired individuals is preferable to one which did not.

Is there a difference between the analysis of the tax treatment of "brown" and "white" ice cream lovers, and the fair haired and dark haired individuals with identical labor supply functions? The question, seemingly, hinges on what are admissible distinctions. But there is, within the utilitarian framework as it has customarily been applied to the analysis of tax structures, no way of determining what are and are not admissible distinctions. If hair color or sex is not admissible, why should a seemingly equally capricious aspect of an individual, the color which he likes his ice cream, be the basis of differentiation? If the preferred color is not an admissible basis, should the preferred flavor -- chocolate versus vanilla -- be admissible? And if the choice among ice creams is not admissible, should the choice between ice cream and cheese be an admissible¹ basis, or the

¹ The fact that the technology for producing two colors of ice cream may be identical, while the technologies for producing cheese and ice cream are different, does not, I think, provide a persuasive basis for differentiation.

choice between goods and leisure? These issues have always been lurking quietly in the background in the analytical discussions of the structure of taxation; the point of the simple models presented in the first two sections of the paper is to examine a context in which they cannot be avoided, to present them in as pristine a form as possible.

There are, at this juncture, three approaches that can be taken.

(1) One can argue that any distinction is admissible; discrimination may well be optimal.

(2) Alternatively, one can define a set of admissible distinctions. This unfortunately removes one of the great advantages of the utilitarian approach to the analysis of tax structures; it presumes the existence of a prior principle for the determination of the set of admissible bases of differentiation. How are we to know, then, that this prior principle should not, at the same time, be reflected in the design of the tax structure itself?

The same criticism can, of course, be levied against the principle of horizontal equity itself. This is usually put as requiring that "people in equal positions should be treated equally..." (Musgrave, 1959). But what is meant by "equal positions": are chocolate and vanilla ice cream lovers in equal positions? Again, it would appear that the critical question is, what are admissible distinctions; the principle of horizontal equity seems to provide little guidance. As Rosen has pointed out (1978), "Customarily, 'equal positions' are defined in terms of some observable index of ability to pay such as income, expenditure, or wealth." But to choose say income or wealth as the basis is virtually equivalent to defining income as the horizontal equitable tax; differential commodity taxes are, by definition, then horizontally inequitable.

This approach attempts to define "equal positions" in terms of some opportunity sets (income or wealth are used as surrogates for "ability to pay.") In contrast, Atkinson and Stiglitz (1976) and Feldstein (1976) have taken a "welfare approach": As Feldstein has put it,

"If two individuals would be equally well off (have the same utility level) in the absence of taxation, they should be equally well off if there is taxation.

[Furthermore,] taxes should not alter the utility ordering."

But this formulation too is not completely persuasive: first, it requires a higher level of cardinality in the utility assessment even than that required by the utilitarian approach. We must be able to compare levels of utility as well as differences. Second, let us consider what this definition implies for our chocolate-vanilla ice cream example. If initially, the chocolate and vanilla ice cream sell at the same price, then differential taxation is horizontally inequitable. But now, assume that the cost of chocolate is increased slightly. The chocolate lover is "disadvantaged." The Feldstein formulation suggests not only that we could not use the tax system to attempt to restore "equality" but that any taxes we impose must result (if it is to be horizontally equitable) in chocolate lovers being worse off than vanilla lovers. For example, assume that the supply elasticity of chocolate is greater than that of vanilla.

A uniform lump sum tax might therefore result, in the new general equilibrium, in the price of chocolate being below that of vanilla. In the Feldstein definition, the uniform lump sum tax (which, given that all individuals have the same endowments, would in the conventional formulation be horizontally equitable) is thus horizontally inequitable.

In either formulation, virtually any tax system will have some degree of horizontal inequity; one needs, then, to trade off horizontal equity with other desiderata of a good tax system. One needs, then, a Meta-principle for evaluating tax systems.¹

3) Finally, we can attempt to retain the utilitarian approach, but argue that the particular formulation in the current optimal tax literature is inadequate. The governments -- the individuals who are in the possession of the power to exercise the power to tax -- are not likely to impose a truly random tax. The existence of differential treatment means that there will be incentives to bribe (in one way or another) those in the power to determine who is to be treated favorably. It is this belief in the corruptability of power which may have provided the motivation for the restriction on differential treatment in the American constitution. These considerations are, I think, relevant in assessing alternative tax codes, e.g. in the desirability of taxing different commodities at the same or different rates.

¹ There are ad hoc approaches defining an index of horizontal inequity and an index of vertical equity, and positing a social welfare function giving tradeoffs between the two. This seems close to assuming what is to be analyzed. The index used by King (1980) seems open to the objections made above.

The dangers of differentiation lie not only in the favorable treatment that may be -- and have been -- obtained by special interest groups. Admitting the possibility that some commodity may be taxed at a higher rate opens up the possibility of using the tax system as an instrument for the (possibly mistaken) public wrath against one industry or another.

Thus a utilitarian, assessing the advantages and disadvantages of alternative tax systems, should take into account how such systems would actually be implemented. It is within such a broader perspective that some form of the principle of horizontal equity may well be consistent with utilitarianism.¹

10. Rawls and Horizontal Equity

Some readers have suggested that there is a close relationship between the principle of horizontal equity and Rawls' principle of justice. And just as Rawls argues that justice takes precedence over other social principles, so too should horizontal equity. Thus, one should not trade-off horizontal equity with other social objectives.

But our analysis has shown that each individual's expected utility may be higher if he is confronted with a random tax structure. Thus, behind the veil of ignorance, each would favor random taxation, if he could be assured that the tax would be truly levied in a random way.

¹ If this view is correct, then the kind of analysis of horizontal equity contained in Feldstein, Rosen and King may not be focusing on the critical issues. Within a general equilibrium context, changes in taxes -- like changes in technology -- affect different individuals very differently. Welfare rankings -- if they could be defined -- may well be reversed. But there is nothing sacred about the pre-tax ranking. What we are concerned with is some notion of arbitrary distinctions.

But the individual may not believe that the tax would be levied in a truly random way. Indeed, aware of the corruptability of government, he might reason that if the Constitution allows differential taxation (of commodities, individuals, etc.) then the political process will result in some individual being advantaged relative to others, simply because of their ability to exercise political power. Thus differentiation would clearly violate the principle of justice. And because it may not be possible to write the Constitution in such a way that it would allow just differentiation, but that it would not allow unjust differentiation, it may be preferable simply to restrict the ability of the government to impose differential taxation.¹

11. Concluding Remarks

This paper has established that, far from being able to derive the principle of horizontal equity from utilitarianism, the principle is actually inconsistent with utilitarianism in a variety of circumstances; most notably, we have derived conditions under which random taxation is optimal. Indeed, there are potentially important economic situations where Pareto optimality and horizontal equity are inconsistent (in both an ex ante and ex post sense). Such inconsistencies force us to re-evaluate our ethical principles: either utilitarianism or the principle

¹ This still leaves a number of questions unresolved: is non-differentiation consistent with all consumption being taxed at the same rate, or with all income being taxed at the same rate; the former is equivalent to a wage tax, and thus implies that interest is exempt from taxation, while the latter is equivalent to a tax on future consumption at a higher rate than on a tax on present consumption.

of horizontal equity -- at least in the conventional forms -- must be abandoned. We have suggested a more general utilitarian approach, within which the two principles may be consistent, but which, at the same time, casts considerable doubt on the optimality of the kinds of tax structures which have been derived within the conventional utilitarian framework.

APPENDIX A

Derivation of curvature of indifference curve (Section 3)

a) To show that at $p^A = p^B$

$$\frac{d^2 p^A}{d(p^B)^2} = -\frac{2}{p} [\rho - \epsilon - 2\eta]$$

we make use of Roy's formula

$$(A.1) \quad V_p = -CV_I$$

Differentiating with respect to I, we obtain

$$(A.2) \quad V_{pI} = -\frac{\partial C}{\partial I} V_I - CV_{II},$$

while differentiating with respect to p we obtain

$$(A.3) \quad V_{pp} = -\left(\frac{\partial C}{\partial p} V_I + CV_{Ip}\right);$$

substituting (A.2) into (A.3), we obtain

$$(A.4) \quad V_{pp} = -\left(\frac{\partial C}{\partial p} V_I - C \frac{\partial C}{\partial I} V_I - C^2 V_{II}\right)$$

Using the Slutsky equation, we have

$$(A.5) \quad V_{pp} = -\left(\frac{\partial C}{\partial p}\right) \frac{V_I}{V} + 2C \frac{\partial C}{\partial I} V_I + C^2 V_{II}$$

which, with some rearrangement, yields (3.7).

APPENDIX B

Analysis of Ex Post Randomization

The first order condition corresponding to the maximization problem (4.1) is

$$(B.1) \quad \frac{U_1[(\omega + \Delta - t)L, L](\omega + \Delta - t) + U_1[(\omega - \Delta - t)L, L](\omega - \Delta - t)}{2} + EU_2 = 0$$

Differentiating W with respect to Δ , we obtain

$$(B.2) \quad \frac{dW}{d\Delta} = \frac{U_1[(\omega + \Delta - t)L, L]L - U_1[(\omega - \Delta - t)L, L]L}{2} + \frac{\partial W}{\partial L} \frac{dL}{d\Delta} - \frac{U_1[(\omega + \Delta - t)L, L]L + U_1[(\omega - \Delta - t)L, L]L}{2} \frac{dt}{d\Delta}$$

To evaluate this, we need to calculate

$$(B.3) \quad \left. \frac{dt}{d\Delta} \right|_{\bar{R}} = - \frac{\bar{R}}{L^2} \frac{dL}{d\Delta} .$$

Since

$$(B.4) \quad \frac{dL}{d\Delta} = \frac{\partial L}{\partial \Delta} + \frac{\partial L}{\partial t} \frac{dt}{d\Delta}$$

$$(B.5) \quad \left. \frac{dt}{d\Delta} \right|_{\bar{R}} = - \frac{\frac{\partial L}{\partial \Delta} \frac{\bar{R}}{L^2}}{1 + \frac{\partial L}{\partial t} \frac{\bar{R}}{L^2}} .$$

Since the denominator is positive for efficient levels of taxation, $\frac{dt}{d\Delta}$ is of opposite sign to $\frac{\partial L}{\partial \Delta}$.

To calculate $\frac{\partial L}{\partial \Delta}$, we differentiate the first order condition (B.1):

$$(B.6) \quad - \frac{\partial L}{\partial \Delta} = \{ U_{11} L(\omega + \Delta - \tau) - U_{11} L(\omega - \Delta - \tau) + U_1((\omega + \Delta - \tau)L, L) \\ - U_1[(\omega - \Delta - \tau)L, L] + U_{21}[(\omega + \Delta - \tau)L, L]L - U_{21}[(\omega - \Delta - \tau)L, L]L \} / \\ \{ U_{11}(\omega + \Delta - \tau)^2 + U_{11}(\omega - \Delta - \tau)^2 + 2U_{21}(\omega + \Delta - \tau) + 2U_{21}(\omega - \Delta - \tau) \\ + 2EU_{22} \} \\ = 0 \quad \text{when } \Delta = 0.$$

At $\Delta = 0$ the first term of (B.2) is obviously zero, the second term is zero, since utility maximization implies

$$(B.7) \quad \frac{\partial W}{\partial L} = 0$$

and the third term is zero since, from (B.5) and (B.6), $\frac{dt}{d\Delta} = 0$.

Similarly,

$$\begin{aligned}
 \frac{d^2 W}{d\Delta^2} &= \frac{U_{11}[(\omega + \Delta - t)L, L]L^2 + U_{11}[(\omega - \Delta + t)L, L]L^2}{2} + \left\{ \frac{dU_1[(\omega + \Delta - t)L, L]L}{dL} \right. \\
 &\quad \left. - \frac{dU_1[(\omega - \Delta - t)L, L]L}{dL} \right\} \frac{dL}{d\Delta} + \frac{\partial^2 W}{\partial L^2} \left(\frac{\partial L}{\partial \Delta} \right)^2 \\
 &\quad - \left\{ U_1[(\omega + \Delta - t)L, L]L^2 - U_1[(\omega - \Delta - t)L, L]L^2 + \right. \\
 &\quad \left. \frac{d\{U_1[(\omega + \Delta - t)L, L]L + U_1[(\omega - \Delta - t)L, L]L\}}{dL} \frac{dL}{d\Delta} \right. \\
 &\quad \left. - (U_{11}[(\omega + \Delta - t)L, L] - U_{11}[(\omega - \Delta - t)L, L])L^2 \right\} \frac{dt}{d\Delta} \\
 &\quad - \frac{U_1[(\omega + \Delta - t)L, L]L + U_1[(\omega - \Delta - t)L, L]L}{2} \frac{d^2 t}{d\Delta^2} \\
 &\quad + \frac{U_{11}[(\omega + \Delta - t)L, L]L^2 + U_{11}[(\omega - \Delta - t)L, L]L^2}{2} \left(\frac{dt}{d\Delta} \right)^2 \\
 &\quad + \frac{\partial W}{\partial L} \frac{d^2 L}{d\Delta^2} + \frac{\partial^2 W}{\partial L \partial \Delta} \frac{dL}{d\Delta} + \frac{\partial^2 W}{\partial L \partial t} \frac{dt}{d\Delta} \frac{dL}{d\Delta}
 \end{aligned}$$

The second and fifth terms are zero at $\Delta = 0$; the third term is zero because of (B.6); the fourth and ninth is zero because of (B.7), the sixth, eighth and eleventh are zero because at $\Delta = 0$, $\frac{dt}{d\Delta} = 0$ (from (B.5) and (B.6)), and the tenth is zero because of (B.3) - (B.6).

REFERENCES

- Arrow, K. J. [1970], Essays in the Theory of Risk Bearing, Chicago: Markham.
- Atkinson, A. B. [1970], "On the Measurement of Inequality," Journal of Economic Theory, pp. 244-263.
- Atkinson, A. B. and J. E. Stiglitz [1972], "The Structure of Indirect Taxation and Economic Efficiency," Journal of Public Economics, Vol. 1, pp. 97-119.
- Atkinson, A. B. and J. E. Stiglitz [1976], "The Design of Tax Structure: Direct Versus Indirect Taxation," Journal of Public Economics, Vol. 6, pp. 55-75.
- Atkinson, A. B. and J. E. Stiglitz [1980], Lectures on Public Finance, New York and London: McGraw-Hill.
- Boiteux, M. [1956], "Sur la Gestion des Monopoles Publics Astreints a l'equilibre budgetaire," Econometrica, Vol. 24, pp. 22-40.
- Dasgupta, P. and J. E. Stiglitz [1980], "Uncertainty, Industrial Structure and the Speed of R&D," Bell Journal of Economics, forthcoming.
- Diamond, P. A. and J. A. Mirrlees [1971], "Optimal Taxation and Public Production," American Economic Review, Vol. 61, pp. 8-27 and 261-278.
- Feldstein, M. [1976], "On the Theory of Tax Reform," Journal of Public Economics, Vol. 6, Nos. 1,2, July/August, pp. 77-104.
- Gilbert, R. and J. E. Stiglitz [1979], "Entry, Equilibrium, and Welfare," mimeo.
- Gordon, R. [1976], "Essays on the Causes and Equitable Treatment of Differences in Earnings and Ability," Ph.D. thesis, M.I.T.
- King, Mervyn A. [1980], "An Index of Inequality: With Applications to Horizontal Equity and Social Mobility," Social Science Research Council Programme: Taxation, Incentives and the Distribution of Income Paper No. 8.
- Lerner, A. P. [1944], The Economics of Control, New York: Macmillan.
- Mirrlees, J. A. [1971], "An Exploration on the Theory of Optimal Income Taxation," Review of Economic Studies, Vol. 38.
- Mirrlees, J. A. [1972], "The Optimum Town," Swedish Journal of Economics, Vol. 74, pp. 114-135.
- Mirrlees, J. A. [1975], "A Pure Theory of Underdeveloped Economies," in L. A. Reynolds (eds.), Agriculture in Development Theory, New Haven: Yale University Press.

Musgrave, R. A. [1959], The Theory of Public Finance, New York: McGraw-Hill.

Musgrave, R. A. [1976], "Optimal Taxation, Equitable Taxation and Second-Best Taxation," Journal of Public Economics, Vol. 6, Nos. 1 and 2, July/August, pp. 3-16.

Newbery, D. [1978], "Stochastic Limit Pricing," Bell Journal of Economics, Vol. 9, pp. 260-269.

Pratt, J. W. [1964], "Risk Aversion in the Small and in the Large," Econometrica, Vol. 32, pp. 122-136.

Ramsey, F. P. [1927], "A Contribution to the Theory of Taxation," Economic Journal, Vol. 37, pp. 47-61.

Rosen, H. S. [1978], "An Approach to the Study of Income, Utility, and Horizontal Equity," Quarterly Journal of Economics, Vol. 92, pp. 307-322.

Rothschild and Stiglitz [1971], "Increasing Risk: II, Its Economic Consequences," Journal of Economic Theory, Vol. 3, March, pp. 61-84.

Salop, S. [1979], "Strategic Entry Deterrence," American Economic Review, Vol. 69, pp. 335-338.

Stiglitz, J. E. and P. S. Dasgupta [1971], "Differential Taxation, Public Goods, and Economic Efficiency," Review of Economic Studies, Vol. 38, pp. 151-174.

Stiglitz, J. E. [1972], "Taxation, Risk Taking and the Allocation of Investment in a Competitive Economy," in M. Jensen (ed.), Studies in the Theory of Capital Markets, pp. 294-361.

Stiglitz, J. E. [1973], "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," presented at Berlin Symposium on Planning, August 1973 (also IMSSS Technical Report No. 166, Stanford University).

Stiglitz, J. E. [1975], "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," Bell Journal of Economics, Vol. 6, pp. 552-579.

Stiglitz, J. E. [1976a], "Utilitarianism and Horizontal Equity: The Case for Random Taxation," IMSSS Technical Report #214, August, Stanford University.

Stiglitz, J. E. [1976b], "The Efficiency Wage Hypothesis, Surplus Labor, and the Distribution of Income in L.D.C.'s," Oxford Economic Papers, forthcoming.

Stiglitz, J. E. [1981], "Self-Selection and Pareto Efficient Taxation,"
Journal of Public Finance, forthcoming.

Stiglitz, J. E. [1982], Information and Economic Analysis, Oxford
University Press.

Weiss, L. W. [1976], "The Desirability of Cheating Incentives and
Randomness in the Optimal Income Tax," Journal of Political Economy,
Vol. 84, pp. 1343-1352.