Taxation and Heterogeneous Preferences

by

Sören Blomquist*
Uppsala University

and

Vidar Christiansen**
University of Oslo

Abstract
Non-linear income taxes and linear commodity taxes are analysed when people differ with respect to ability, high-skilled agents have heterogeneous preferences, and neither individual abilities nor preferences are observable. A taxonomy of preferences is presented distinguishing between preferences for market goods on the one hand and the trade-off between income and leisure on the other. Even if the focus is on the latter, it is also a purpose to highlight the interaction between preferences for market commodities and the income-leisure trade-off and the implications for tax policy.

Pure income tax optima, which may be bunching or separating optima, are characterised. In particular the income tax may not be able to distinguish between those low-income people who have low ability and those who have strong preference for leisure. As is shown there may still be an impact on the optimum income tax schedule as it will depend on the composition of the population with respect to types of individuals. Finally, the paper addresses what can be achieved by commodity taxes when preferences are heterogeneous.

Keywords: Optimum Taxation, Heterogeneous Preferences, Asymmetric Information

JEL Numbers: H21, H23

* Department of Economics, Uppsala University, Box 513, 751 20 Uppsala, Sweden and the Institute for Futures Studies. E-mail: Soren.Blomquist@nek.uu.se
** Department of Economics, University of Oslo, P.O.Box 1095 Blindern, 0317 Oslo, Norway, e-mail: vidar.christiansen@ econ.uio.no
1. Introduction.
There can be little doubt that in practice people have heterogeneous preferences and therefore choose different consumption bundles even if they have the same income or wage rate and face the same prices. Some have a strong taste for travelling, dining at restaurants, and going to the theatre, whilst others prefer to invest in housing and household durables to enjoy more time at home. Some prefer to put a lot of effort into a highly paid job, whilst others prefer a quiet and less stressful life. Although it is easy to recognise the variety of tastes in the population, optimum tax theory has been surprisingly silent on this score. Whilst a standard assumption is that various agents have different endowments or skill levels that motivate a distributional role for taxes, it has almost invariably been assumed that the same utility function motivates individual choices. An important question is whether neglecting differences in preferences is justified or whether the recognition of heterogeneity of preferences should impact the choice of tax policy. Should the income tax schedule be responsive to the distribution of preferences? Does preference heterogeneity make a difference to the case for commodity taxes?

To address these questions we shall consider a population of individuals differing along two dimensions. There are low-skilled and high-skilled individuals with the latter group being comprised of people with strong preference for leisure and people with weak preference for leisure. All low-skilled are assumed to have weak preference for leisure.

Roughly speaking there are two approaches to the issue of how to treat equally skilful people with different preferences. One is what we may call the horizontal equity approach, which imposes the restriction that equally skilful individuals should be treated in some sense equally, for instance by being faced with the same tax liability irrespective of preferences. The other one is the social welfare approach, which implies that a social welfare function for all individuals, with their respective skills and preferences, is maximised even if that may imply different treatment of people with identical skill levels. The discussion of how to treat different types of people is often taking place under the explicit or implicit assumption that there is full information about individual skills and preferences. It is then conceivable that one may impose a high, and possibly uniform, tax on high-skilled individuals irrespective of preferences, and tax leniently, or make a transfer to, low-skilled individuals.
When there is asymmetric information about individual skills and preferences the perspective may be different. Even if at a no-tax point of departure there is considered to be no case for redistribution between equally skilful people, the case for equal treatment may be weakened as soon as vertical redistribution takes place, and asymmetric information constrains how taxes can be imposed. Assume that one starts redistributing from high-skilled people to low skilled people. In the extreme case high-skilled people with strong preference for leisure and low-skilled individuals will make the same choice of observable income points. Any income point intended for the low-skilled will be adopted by the high-skilled with strong preference for leisure, and the equal treatment of all high-skilled breaks down. An attempt to redistribute from all high-skilled to low-skilled is not be feasible as the high-skilled with strong preference for leisure will in fact be treated as low-skilled. One cannot at the same time treat all high-skilled equally and high-skilled and low-skilled differently. Vertical redistribution and horizontal equity becomes irreconcilable objectives.

In less extreme cases different income points may be assigned to the high-skilled with strong preferences for leisure and low-skilled individuals, respectively, but only subject to a self-selection constraint. Attempted redistribution from high-skilled to low-skilled may induce high-skilled with strong preference for leisure to start mimicking the low-skilled. The income point assigned to the low-skilled must be sufficiently distorted to discourage the high-skilled with strong preference for leisure to mimic by choosing the same income point. The implication is that the transfers to the low-skilled are costly in terms of distortions. However, the cost can be reduced by lowering the transfer from the high-skilled with strong preference for leisure who is the potential mimicker, whilst relying more on transfers from the high-skilled with weak preference for leisure. This will alleviate the cost of transfers to the low-skilled but only by abandoning equal treatment of the high-skilled with different preferences (horizontal equity). The general argument is that transfers from different types of high-skilled individuals may imply different costs under asymmetric information and hence employing the less costly alternative may be preferable, but only by violating horizontal equity.

Our conclusion is that it may be crucial to distinguish between full information and asymmetric information regimes when it comes to discussing the tax treatment of equally skilful individuals. It may be that horizontal equity can be enforced without serious problems under full information, or there may be no social welfare case for
redistribution between different high-skilled persons under full information, but this
may be entirely different under asymmetric information. Imposing horizontal equity
may then be costly. A social welfare approach may treat the various types of high-
skilled individuals differently, not because that is desirable per se, but because it is
part of a scheme involving transfers to low-skilled individuals. Assuming that we do
indeed treat the high-skilled individuals differently by taxing more leniently those
with strong preference for leisure (in order to reduce the cost of vertical
redistribution), it is plausible that different social marginal utilities of income will be
assigned to the high-skilled individuals even if no such discrepancy would exist in the
no-tax regime. Below we shall pursue the welfare approach to elaborate on
conceivable income tax optima when high-skilled individuals may be taxed
differently.

Heterogeneity of preferences can, roughly classified, be of two kinds. People
can have different trade-offs between income and leisure, or they can have different
preferences for consumption bundles bought in the market, in the sense that they will
buy different consumption bundles even if they have the same income and leisure and
face the same prices. We may note that various combinations of the two types of
preferences are conceivable. People may be endowed with different preferences for
leisure, while still having the same preferences for market goods. People may also
have different preferences for leisure because they have different preferences for
market goods. The reason is simple. Since income is used for buying market goods,
the preferences for market goods may affect the trade-off between income and leisure.
We shall survey preference heterogeneity in more detail below.

As has been discussed by some authors (e.g. Cuff (2000)), heterogeneity of
preferences can be given different interpretations. A stronger preference for a good
may sometimes be interpreted as meaning that a person has a stronger need for the
good in question. Indeed the person may have a disadvantage or handicap that can
only be alleviated by consuming some good. In such a case we will perceive this
person as being worse off if he does not get more than others of the good he has a
special need for, and even then he may be worse off as he may never be fully
compensated for his handicap. A clearcut example may be a person in need of
medication. On the other hand it is also conceivable that a sort of handicap may
induce a weaker preference for a commodity. An illiterate person will have a weak
preference for books. A person less able to enjoy music will have a weaker preference
for music records or concert tickets. Conversely a person who is more well endowed in the relevant respects will have stronger preferences for these goods. Thus a strong preference for a good may be interpreted either as reflecting a handicap or a talent (a particular endowment or ability to enjoy a good) depending on the underlying nature of goods and consumers.

We will consider a heterogeneity of preferences that neither reflects a superior nor an inferior overall endowment. People are simply different, without being systematically more or less well endowed physically, mentally or emotionally. They have different pure tastes. A central assumption is that preferences are not observable. The government knows the distribution of preferences, but not the identity of people with the respective tastes.

The very small literature on taxes on a population of agents with heterogeneous preferences has adopted various approaches. Sandmo (1993) asked how one should tax people who are endowed with the same resources, but have different preferences. Even in this very simplified setting there are few general conclusions to be drawn. Tarkiainen and Tuomala (1999) used numerical specifications to derive computationally optimum non-linear tax schedules for a continuum of agents simultaneously distributed by skill level and preferences for income and leisure. Their findings are interesting as illustrations of possible optima. However, their results could be sensitive to the specifications that have been made. Assuming a special class of utility functions, Boadway et al. (2002) considered a discrete distribution of individuals assuming there are two skill levels and two types of preferences. However the high-skilled person with strong preference for leisure cannot be distinguished from the low-skilled person with weak preference for leisure and must always be treated equally. The aim of their analysis is to trace the income tax implications of assigning different welfare weights to the high-skilled person with strong preference for leisure. Saez (2002) discusses the role of commodity taxes supplementing an income tax in a very general framework with a continuum of individuals from a multidimensional set. The focus of his study is to identify conditions under which the Atkinson-Stiglitz result of zero commodity taxes can be recovered in a setting with heterogenous preferences. In contrast the present paper uses a model with a small, discrete number of types of individuals, and distinguishes

---

1 A main finding is that preference heterogeneity does not necessarily lower the progressivity of the income tax.
more clearly between people with different skills and people with different preferences. In contrast to Saez we also provide the formulas characterizing the optimal commodity taxes. Jordahl and Micheletto (2002) discussed tax policy implications of imposing a horizontal equity requirement when preferences for leisure vary. Cuff (2000) considered heterogeneous preferences in the context of workfare adopting a special class of utility functions.

In view of the existing literature, and in particular the discussion of Sandmo (1993) an appropriate question seems to be: Why is it so hard to draw conclusions about distributional policy when preferences vary? Part of the answer seems to be that making interpersonal utility comparisons is particularly hard when preferences are acknowledged to be heterogeneous. Even if in principle we find ourselves able to compare utilities, who is judged to be better off may depend on the parameters of the economy such as prices and wage rates. Even if person I is judged to be better off than person II for one set of parameters the difference may be evened out or reversed if there is a change to a different set of parameters. The implication is that even if we are willing to base utility comparisons on strong assumptions there is the further difficulty of obtaining sufficiently robust assumptions, i.e. assumptions that are sufficiently invariant to the specific parameters of the economy we study. However, there is even a further challenge. As emphasised by Sandmo (1993) the direction of redistribution should be governed by differences in marginal rather than total utilities. This would be a rather uninteresting statement if there were a unique relationship between total and marginal utilities, but this will only be true under very restrictive assumptions. In general the marginal utility of income does not only depend on which indifference curve that is assigned to the person but also on the particular commodity bundle that generates the corresponding utility.

Throughout our discussion we shall make the assumption that, based on pure (marginal) utility comparisons it would be desirable to make further transfers to the unskilled individuals. The actual transfer is, however, limited by the informational constraints facing the decision-maker. This is in line with the standard assumption in the (conventional) tax theory with individuals of different skills, but homogeneous preferences. In our context we shall even assume that the desirability of improving the welfare of the low-skilled is valid when comparing with both high-skilled types. This implies a comparison across preferences and is a stronger assumption than the conventional one. However, we shall leave open the outcome of (marginal) utility
comparisons between the high-skilled individuals with different preferences, and consider alternative cases.

Heterogeneity of preferences can affect the choice of income tax schedule, the case for employing commodity taxes and the rates of these taxes. We shall address all these aspects.

In the next section we provide a taxonomy of preference heterogeneity. Section 3 addresses various pure income tax optima when preferences are heterogeneous, and examines the effects of changing the composition of the population. Section 4 discusses what commodity taxes can achieve when preferences are heterogeneous. In section 5 we take a brief, closer look at some of the underlying problems of interpersonal utility comparisons largely bypassed in the preceding sections. Section 6 concludes.

2. A taxonomy of preference heterogeneity.

In general people have preferences over bundles of market commodities and labour, and these preferences may be identical, or they may be heterogeneous in various ways. We may in general write the utility function as \( u^i(x, h) \) where the subscript \( i \) indicates a particular type of preference, \( x \) denotes a vector of market commodities, and \( h \) is hours of work. Since there is a unique relationship between labour and leisure, we may choose to express preferences in terms of either of these variables. We opt for labour as argument in the utility function.

A problem with this general approach is that it covers a multitude of preferences, and without further restrictions almost any behavioural variation and any welfare considerations are possible. Therefore it is helpful to start by identifying certain classes of preferences, distinguishing between preferences for market goods on the one hand and the trade-off between market goods and leisure on the other hand.

Let us consider the following categories of preferences:

i) There are identical preferences for market commodities, but different preferences for labour, which is assumed to be weakly separable from market goods in the utility function. The utility function belongs to the class \( u^i(x, h) = u^i(f(x), h) \). Choosing his bundle of market commodities, the consumer will maximise \( f(x) \) s.t. the budget constraint \( B = px \), with p
denoting the price vector, and we get the maximum value function \( f^*(p, B) \).

The preferences for income and labour are then expressed by

\[
v^i(p, B, h) = u^i(f^*(p, B), h).
\]

ii) Labour is not weakly separable from market goods, but conditional on labor there are uniform preferences for market goods, while preferences for labour differ. The utility function is of the form \( u^i(x, h) = u^i(f(x, h), h) \), assumed to be increasing in the first argument and deceasing in the second. The preferences for income and labour are expressed by

\[
v^i(p, B, h) = u^i(f^*(p, B, h), h).
\]

iii) Labour is weakly separable from market commodities. Preferences for market goods are heterogeneous, whilst people have the same preferences for leisure and market commodity aggregates. The utility function is of the form \( u^i(x, h) = u(f^i(x), h) \), where the interpretation is that it is the ordinal properties of the \( f \)-function that may vary between consumers. A specific example might be \( x_1^{a_i} x_2^{b} - h^b \) where \( a_i \) and \( b \) are positive parameters and \( a_i \) is type specific. The preferences for income and labour are expressed by

\[
v^i(p, B, h) = u(f^i*(p, B), h).
\]

iv) Labour is not weakly separable from market commodities, and preferences for market commodities are heterogeneous. Obviously no separability is obtained in the corresponding \( v^i \)-function.

On the basis of this discussion we recognise that the income/labour trade-off expressed by a utility function \( v(B, h) \) can originate from several underlying basic preference patterns from which its properties are derived. It is helpful to distinguish between different preferences for commodities (as in iii and iv) and different preferences for labour (as in i and ii) as sources of heterogeneity of preferences for \( B, h \)-bundles. In particular, we note how (non-)separability between disposable income and labour depends on the properties of the underlying basic utility function.

Working with \( v(B, h) \) it is often helpful to impose restrictions that simplify the analysis. Probably the most popular class of functions being deployed is the additively...
separable form \( v(B,h) = \varphi(B) - g(h) \). It is therefore of interest to explore how this class relates to the underlying preferences. Considering the general form of preferences for market commodities, \( f(x,h) \), the maximum value function \( f^* \) is derived by maximising \( f(x,h) \) s.t. the budget constraint \( \sum_i p_i x_i = B \), and we get the familiar first order conditions \( f_i = \lambda p_i \), where \( f_i \) is the partial derivative w.r.t. \( x_i \) and \( \lambda \) is the Lagrange multiplier. Suppressing \( p \) we get \( f^*(B,h) = f(x_1(B,h),...x_i(B,h),...,h) \) and \( \frac{\partial f^*}{\partial h} = \sum_i f^*_i \frac{\partial x_i}{\partial h} + f^*_h = \lambda \sum_i p_i \frac{\partial x_i}{\partial h} + f^*_h = f^*_h \) due to the budget constraint.

Since \( h \) would have no effect on preferences for market goods if \( f_h \) were zero, we recognise that the maximum value of \( f \) will be affected by \( h \) and \( v(B,h) \) cannot be written as an additively separable utility function. We can conclude that if the disposable income \( B \) is used to buy several commodities a necessary condition for additively separable preferences for income and labour, \( V(B,h) = \varphi(B) - g(h) \), is that the basic utility function is also additively separable of the form \( u(x,h) = f(x) - g(h) \). We recognise that case ii) above is not consistent with an additively separable \( V \)-function.³

Most previous contributions have abstracted from the plurality of market commodities and have just addressed preferences for disposable income and labour without discussing how the disposable income is spent. We may note that when there are underlying preferences for multiple commodities the corresponding price vector affects the location and shape of indifference curves in \( h,B \)-space (and \( Y,B \)-space). We shall come back to this issue below.

### 3. Income taxation when there is heterogeneity in preferences between leisure and consumption

We will apply a three-type version of a Mirrlees (1971) income tax model, which is similar to the extensively employed two-type model first introduced by Stern (1982) and Stiglitz (1982). Each type of person has a skill level, which is reflected by his (exogenous) wage rate \( (w) \). We will index the three types of individuals of our model

---

2 Further popular, special cases are the quasi-linear functions \( \beta B - g(h) \) or \( v(B) - \gamma h \).
3 If there are only two commodities, a single market commodity and leisure, additive separability imposes restrictions on the indifference map as discussed by (1). With several commodities similar restrictions are imposed on the indifference map for labour and the consumption aggregate \( f(x) \).
economy in the following way: type 1 is a low-skilled person with weak preference for leisure, type 2 is a high-skilled person with strong preference for leisure and type 3 is a high-skilled person with weak preference for leisure. Before tax income is defined as $Y = wh$. We will denote the wage rate for a low skill by $w^1$ and for a high skill by $w^3$. The three types will have the before tax incomes $Y^1 = w^1h^1$, $Y^2 = w^3h^2$ and $Y^3 = w^3h^3$. Even if these statistical properties of the population are public information, we shall assume that there is asymmetric information in the sense that the government (tax authority) is informed neither about individual skill levels nor individual preferences.

As was the conclusion of Sandmo (1993) we cannot in general tell whether we would like to redistribute income between equally skilful individuals. Accepting this fact, we shall assume that there is indeed no case for such redistribution. We want to argue that distribution among equally skilful people should not be considered in isolation, but rather in the wider context of redistribution across skill levels. This leads us to the conventional assumption that it is desirable to make transfers from (both types of) high-skilled to lowed-skilled individuals. From this perspective it turns out the treatment of equally skilful people tends to be governed by the costs of making transfers to the low-skilled from the respective types of high-skilled individuals, as we shall see in further detail below. The implication is that even if there is no presumption that income should be redistributed across people with heterogeneous preferences, differential treatment of equally skilful people may be an implication of pursuing vertical redistribution.

In the absence of any initial basis for differential treatment we shall assume that the same tax is imposed on both types of high-skilled individuals. The objective is to appropriate resources for the benefit of the low-skilled. This can happen in two ways as the high-skilled may forego consumption or leisure, and the responses will depend on the respective preferences. A person reluctant to forego consumption will increase his labour supply. We cannot in general judge whether the tax liability is a heavier burden for one type than the other and our presumption is that the same tax is imposed as long as redistribution from either person is not subject to any constraint. In this sense horizontal equity is assumed to prevail.
At some point information constraints will start binding as taxes must be designed subject to the self-selection constraints ensuring that each person selects the income point intended for his type. Given the usual monotonicity assumption a person of type 3 will have a flatter indifference curve in $Y, B$ space than a type 1 or a type 2 person. It is flatter than that of type 1 due to the wage rate difference and flatter than that of type 2 because of the different preferences. Depending on the exact form of the preferences a type 2 person may have a flatter or steeper indifference curve through a given point in $Y, B$ space than a type 1 person. In either case we assume the single crossing property holds. There can be two types of optima. There can be a screening optimum where the three types of individuals are assigned three distinct income tax points or a bunching optimum such that the type 1 and 2 persons are assigned the same income point. We characterize a screening optimum in section 3.1 and a bunching case in section 3.2.

As income is being transferred a number of cases may arise depending on the indifference maps.

i. A type 2 person may be the first to start mimicking type 1. This may happen with type 2 having steeper or flatter indifference curves than type 1. The condition that type 2 should not mimic type 1 may be the (only) binding self-selection constraint.

ii. As there further redistribution takes place it may be that also the condition that type 3 should not mimic type 2 (or 2 should not mimic 3) becomes binding.

iii. The self-selection constraint that type 3 should not mimic type 1 may be binding.

3.1 Screening optima
We will initially assume that the preferences are such that the type 1 persons have the steeper slope. This implies that a type 3 person might mimic a type 2 person, by choosing the income point intended for the latter, and a type 2 person might mimic a type 1. There are $n^i$ persons of type $i$. The optimisation problem to be addressed is the
maximisation of an additive welfare function (sum of utilities) subject to the budget and self-selection constraints. The corresponding Lagrange expression is

\[ \Lambda = n^1V^1(B^1, Y^1) + n^2V^2(B^2, Y^2) + n^3V^3(B^3, Y^3) + \alpha^3 \left( V^3(B^3, Y^3) - V^3(B^2, Y^2) \right) + \alpha^2 \left( V^2(B^2, Y^2) - V^2(B^1, Y^1) \right) + \mu(n^1(Y^1 - B^1) + n^2(Y^2 - B^2) + n^3((Y^3 - B^3)) \right) \tag{1} \]

where the Lagrange multiplier \( \alpha^3 \) has been assigned to the self-selection constraint that type 3 should not mimic type 2, and \( \alpha^2 \) is assigned to the self-selection constraint that the high-skilled type with strong preference for leisure should not be mimicking the low-skilled. The last restriction is the budget constraint.

The first order conditions are

\[ B^1: \quad n^1V^1_b - \alpha^2V^2_b - \mu n^1 = 0 \tag{2} \]

\[ Y^1: \quad n^1V^1_y - \alpha^2V^2_y + \mu n^1 = 0 \tag{3} \]

\[ B^2: \quad n^2V^2_b - \alpha^3V^3_b + \alpha^2V^2_b - \mu n^2 = 0 \tag{4} \]

\[ Y^2: \quad n^2V^2_y - \alpha^3V^3_y + \alpha^2V^2_y + \mu n^2 = 0 \tag{5} \]

\[ B^3: \quad n^3V^3_b + \alpha^3V^3_b - \mu n^3 = 0 \tag{6} \]

\[ Y^3: \quad n^3V^3_y + \alpha^3V^3_y + \mu n^3 = 0 \tag{7} \]

Double superscripts indicate that a person, indicated by the former superscript, mimics some other person, indicated by the latter superscript. We denote the income tax function by \( T(Y) \). It follows from eqs. (6) and (7) that the type three persons are undistorted facing a zero marginal tax \( T'(Y^3) = 0 \). For a low skill person we obtain the usual result of a positive marginal tax. Manipulations of eqs. (2) and (3) yield

\[ T'(Y^i) = \rho_1(MRS^1 - MRS^{21}) > 0, \] where the marginal rate of substitution

\[ \rho_1(MRS^1 - MRS^{21}) > 0, \] where the marginal rate of substitution
\[ MRS = -\frac{V_y}{V_B} \] and \( \rho \) denotes a positive term\(^4\). For the high skill type with a strong preference for leisure, the type \( 2 \) persons, manipulations of eqs. (4) and (5) yield a similar expression \( T'(Y^2) = \rho_2(MRS^2 - MRS^{32}) > 0 \). In general we cannot tell which type has the highest marginal utility of income or the higher marginal income tax.

Let us now consider the special case in which the constraint that type three should not mimic type two is non-binding, whilst the other constraint remains effective. This implies that \( \alpha_2 > 0 \) and \( \alpha_3 = 0 \). The first order conditions imply

\[ -V_y^2 / V_B^2 = -V_y^3 / V_B^3 = 1 \quad (8) \]

and

\[ -\frac{V_y^1}{V_B^1} - \frac{V_y^{21}}{V_B^{21}} = \left(1 - \frac{V_y^1}{V_B^1}\right) \frac{\mu n^1}{\alpha^2 V_B^{21}} > 0 \quad (9) \]

It is then easy to recognize that both high-skilled types will be undistorted, while the low-skilled type gets a distorted consumption bundle.

Simple manipulations of the conditions above then imply that

\[ V_B^3 = \mu \quad (10) \]

\[ V_B^1 = \mu + \frac{\alpha^2 V_B^{21}}{n^1} > V_B^3 \quad (11) \]

\[ V_B^2 = \mu \frac{n^2}{n^2 + \alpha^2} = V_B^3 \frac{n^2}{n^2 + \alpha^2} < V_B^3 \quad (12) \]

and accordingly

\[ V_B^1 > V_B^3 > V_B^2 \quad (13) \]

Analogously we find when considering redistribution in terms of leisure (labour).

\[ -V_y^3 = \mu \quad (14) \]

\[ -\frac{V_y^1}{V_B^1} = \mu + \frac{-\alpha^2 V_y^{21}}{n^1} > -V_y^3 \quad (15) \]

\[^4\text{We adopt the standard approach of equating the marginal rate of substitution to the marginal income net of tax, i.e. } -\frac{V_y}{V_B} = 1 - T'(Y) \text{ by which the marginal tax rate is implicitly defined.}\]
\[-V^2_\gamma = \mu - \frac{\eta^2}{n^2 + \alpha^2} < -V^3_\gamma, \quad (16)\]

and accordingly
\[-V^1_\gamma > -V^3_\gamma > -V^2_\gamma. \quad (17)\]

We have assumed that transfers in favour of the low-skilled type are desirable to such an extent that he may be mimicked by high-skilled individuals. If only the mimicking constraint on the high-skilled type with strong preference for leisure is binding, the optimum is characterised by the social marginal utility of income being highest for the low-skilled and lowest for the high-skilled with strong preference for leisure.

It follows from the difference between marginal utilities that ideally one would like to redistribute from type 2 to type 3. The reason why this imbalance persists at the optimum is that, if one were to tax type 2 harder, he would escape by mimicking type 1. Then there would be two options. i. One would have to worsen the situation also for type 1, which implies that a pure transfer from type 2 to type 3 is not feasible, and it is this additional cost in terms of lower welfare for the low-skilled that prevents the redistribution from taking place. ii. Alternatively one could maintain the utility of type 1, but only at the cost of further distorting the allocation, which would imply achieving the same utility at a higher resource cost. There would be a ‘leakage’ implying that for each income unit foregone by type 2, less than one unit would accrue to type 3.

Ideally one would also like to make further transfers from the high-skilled to the low-skilled. A further transfer from the high-skilled with stronger preference for leisure is blocked by the standard mimicking problem. Appropriating a unit of income from the high-skilled does not allow a one unit transfer to the low-skilled as a further distortion is required in order to rule out mimicking.

Let us consider how this regime may arise. Assume that one redistributes from high-skilled to low-skilled until at some point type two starts mimicking type one, and further transfers from two to one becomes more costly as it requires a distortion of type one’s behaviour. This is a case for giving priority to transfers from type three, who is not (yet) induced to mimic. The larger transfers from type three may imply that at some point he is considered to be over-taxed compared to type two in the sense that the income distribution between two and three is skewed in favour of type two. We would then have the situation just described.
We next focus on the case where only self-selection constraint on type three is binding, i.e. that the industrious high skill person should not mimick the low skill. We let $\beta^3$ denote the lagrangian for this constraint. The first order conditions imply

$$-\frac{V^2_Y}{V^2_B} = -\frac{V^3_Y}{V^3_B} = 1$$

and

$$\frac{V^1_Y}{V^1_B} = \frac{-V^3_Y}{V^3_B} = \left(1 - \frac{-V^1_Y}{V^1_B}\right) \frac{\mu n^1}{\beta^3 V^3_B} > 0,$$

where the sign follows from the standard assumption that the high-skilled type has the steeper indifference curve through any given point in $Y,B$-space. We get the conventional situation that the low-skilled person is distorted by a positive marginal tax rate at the optimum (efforts to increase gross income by one unit are rewarded by a smaller amount). Assuming no mimicking between high-skilled individuals, both high-skilled types are undistorted.

Now turning to distributional concerns, we note the further implications

$$V^1_B - V^3_B = \frac{\beta^3 V^3_B}{n^1} + \frac{\beta^3 V^3_B}{n^3} > 0,$$

$$V^1_B = V^2_B + \frac{\beta^3 V^3_B}{n^1} > V^2_B,$$

$$V^2_B = \mu = \frac{n^3 + \beta^3}{n^3} V^3_B > V^3_B,$$

Then $V^1_B > V^2_B > V^3_B$. (23)

and similarly

$$-V^1_Y > -V^2_Y > -V^3_Y.$$ (24)
When only the mimicking constraint on the high-skilled type with weak preference for leisure is binding (i.e. type three is on the verge of mimicking), the optimum is characterised by the social marginal utility of income being highest for the low-skilled and lowest for the high-skilled with strong preference for leisure. In this case the difference of marginal utilities implies that ideally one would like to redistribute from type 3 to type 2. Once again it is the mimicking constraint that prevents such a transfer as it could not take place without adversely affecting the low-skilled person or adding to the resources required to ensure him the same utility as before. Also transfers from high-skilled to low-skilled would to some extent be eroded by mounting self-selection problems. For instance a transfer from type 2 to type 1 would require also ‘bribing’ type 3 or further distorting the transfer to person 1 to avert mimicking.

Briefly considering how this regime might come about, assume that income is transferred from high-skilled to low skilled until at some point type three is induced to mimic type one, whereas type two is not induced to mimic. The latter has such a strong preference for leisure that he will choose a low income/high leisure bundle that will be considered inferior according to the preferences of the low-skilled. Redistribution to the low-skilled is then cheaper when transfers are made from type two rather than from type three since the latter will require a distortion of the behaviour of the low-skilled. This may shift the emphasis to transfers from type three implying an income distribution between two and three in favour of type three, and we get the situation discussed above.

In both cases of a single binding self-selection constraint we have obtained an ordering of marginal utilities. Nevertheless, no straightforward inference about utility levels is available as in general there is no simple relationship between total and marginal utilities. In order to clarify the scope for drawing conclusions about utility levels we shall make a few assumptions, whilst postponing further discussion of the underlying utility comparisons.

In general we can identify three determinants of the social marginal utility of income:

i. Preferences
ii. Utility level
iii. Location on the indifference curve
Assuming that the social planner is inequality averse the social marginal utilities assigned to various individuals will depend on the utility levels that the individuals are considered to obtain. If person A is assumed to be less well off than another person B, inequality aversion implies a case for assigning a higher marginal social utility to the former. However, we cannot conclude that a less well off person will always be given the higher marginal social utility of income as it will also in general change along an indifference curve. In addition it seems plausible that a person with a stronger preference for income will in some sense other things equal derive more utility from additional income than a person with a weaker preference for income. Let us now discuss these factors in turn.

To discuss the preference factor, let us now assume uniform preferences for market commodities at least conditional on labour supply, whilst preferences for income and leisure differ. In order to isolate the preference effect we define other things equal by assuming that two persons with different preferences have the same utility level and the same consumption bundle. We will then make the following assumption

**Assumption U:** When utility levels are the same, and different persons have the same consumption bundle, an individual with stronger preference for income derives more utility from additional income and an individual with stronger preference for leisure derives more utility from additional leisure.

It is sometimes helpful to consider a special class of utility functions.

**Definition (B-quasi-linearity):** A utility function is B-quasi-linear if it belongs to the class \( V = \psi (B-g(h))\), where \( \psi \) and \( g \) are monotonic, increasing functions.

It is called B-quasi-linear because the ordinal preference function can be expressed as a function, \( B-g(h) \), that is linear in B. This class of functions exhibits the property that the social marginal utility of income is constant along an indifference curve as we see that: \( V'_b = \psi'(B-g(h)) = \psi'(\psi^{-1}(V)) \). In general the social marginal utility will depend on the utility level as well as the particular location on an indifference curve.
Analogously we can state

**Definition (h-quasi-linearity):** A utility function is h-quasi-linear if it belongs to the class $V = \varphi(\xi(B) - h)$, where $\varphi$ and $\xi$ are monotonic, increasing functions.

Within this class of functions the marginal utility of leisure is constant along an indifference curve.

We assume inequality aversion in the following sense.

**Assumption I:** The government is inequality averse in the sense that for given ordinal preferences the social marginal utilities of income and leisure decrease if there is a partial increase in disposable income or leisure. In the case of B-quasi-linear preferences the social marginal utility of income is inversely related to the utility level. In the case of h-quasi-linear preferences the social marginal utility of leisure is inversely related to the utility level.

Combining Assumption U and the properties of B-quasi-linearity we can state

**Assumption UB:** When preferences are B-quasi-linear and utility levels are the same, the type with stronger (weaker) preference for income derives more (less) utility from additional income.

Our assumptions and the properties of the B-quasi-linear utility function also imply:

**Lemma 1:** If preferences are B-quasi-linear, and the type of person with stronger preference for leisure (weaker preference for income) has the higher marginal utility of income, he has the lower utility.

This is straightforward. If utility levels were the same, the person with stronger preference for leisure would have a lower marginal utility of income according to assumption UB. Since by Assumption I the marginal utility of income is decreasing in utility, it is only if the individual with stronger preference for leisure has a sufficiently lower utility that he will have the higher marginal utility of income.
Analogously we can state

**Assumption Uh:** When preferences are h-quasi linear and utility levels are the same, the type with stronger preference for leisure derives more utility from additional leisure.

Our assumptions and the properties of the B-quasi-linear utility function also imply:

**Lemma 2:** If preferences are h-quasi-linear, and the type of person with stronger preference for income (weaker preference for leisure) has the higher marginal utility of leisure, he has the lower utility.

The following propositions summarise the characterisations of the two regimes with a single binding self-selection constraint.

**Proposition 1:** When the only binding self-selection constraint is that the high-skilled individuals with stronger preference for leisure (type 2) do not mimic the low-skilled individuals (type 1),

a. both high-skilled individuals are undistorted while the low-skilled face a positive marginal tax rate,

b. the low-skilled individuals have the highest social marginal utility of income, and the high-skilled individuals have the lowest marginal utility of income.

c. under assumptions $I$ and $Uh$ the high-skilled individuals with stronger preference for income (weaker preference for leisure) have a lower utility than the high-skilled with stronger preference for leisure.

We note that part c follows from Lemma 1.

**Proposition 2:** When the only binding self-selection constraint is that the high-skilled individuals with stronger preference for income (type 3) do not mimic the low-skilled persons (type 1),

a. both high-skilled individuals are undistorted while the low-skilled face a positive marginal tax rate,
b. the low-skilled individuals have the highest social marginal utility of income, and
the high-skilled individuals with stronger preference for income have the
lowest social marginal utility of income ($V_B^1 > V_B^2 > V_B^3$).

c. under assumptions I and UB the high-skilled with stronger preference for
leisure have lower utility than the high-skilled with stronger preference for income.

We note that part c follows from Lemma 2.

3.2 The bunching case and comparative statics.
High-skilled people with strong preference for leisure may be hard to distinguish from
low-skilled people. As a polar case we may consider the situation in which type 1 and
type 2 have identical indifference maps in $Y,B$-space. Let $Y_i,B_i$ be the (gross and net) income bundle of type $i$. Since type 1 and
type 2 have identical indifference curves their income points must be on the same
indifference curve. Otherwise, one of the income points would be strictly preferred by
both types, and the other point could not be imposed through self-selection. Moreover,
the same income on the common indifference curve must be assigned to both types at
the optimum. If not, there would always be a gain from switching one of the
individuals to the less resource consuming of the two points. Hence we can set
$Y^2 = Y^1$ and $B^2 = B^1$. Since an income tax cannot discriminate between types 1 and 2, the only interesting tax issue is how the existence of different preferences will
affect the respective taxes on category 3 on the one hand and on categories 1 and 2 on
the other.

We shall confine our attention to the information-constrained optima where
the choice of policy is strictly constrained by the requirement that the high-skilled
type with weak preference for leisure prefers the income point $Y^3, B^3$ actually
intended for him.

It is helpful to introduce the expenditure function $e^i(Y,V)$ expressing the
disposable income required by a person of type $i$ in order to achieve the utility level $V$.

---

5 A special case would be the one presented by Boadway et al (2002).
when his gross income is $Y$. We can describe the economy by the following set of equations.

\[ N(Y^3 - B^3) + M(Y^1 - B^1) = 0 \]  
\[ B^1 = e^1(Y^1, V^1) \]  
\[ B^3 = e^3(Y^3, V^3) \]  
\[ B^1 = e^3(Y^1, V^3) \]

(25) is the budget constraint with $N$ being the number of type three people and $M$ the number of people of type one and type two. (26) and (27) follow from the definition of the expenditure function, and (28) is the strictly binding self-selection constraint that the high-skilled person with weak preference for leisure obtain no higher utility when mimicking (choosing $Y^1, B^1$) than the level $V^3$ achieved when choosing the income point actually intended for him. We approach the social optimisation in two steps, first considering the Pareto optimality and then the welfare maximisation.

Assigning any indifference curve (defined by $V^1$) to type 1 and 2, Pareto optimality is achieved by maximising the utility of type 3 s.t. the budget and self-selection constraint and also pegging the value of $V^1$. We may note that the common indifference curve of type 1 and type 2 is steeper than that of type 3 through any given point in $Y, B$-space. That of type 2 is steeper by the difference in preference for leisure. That of type 1 is steeper by standard arguments (see Stiglitz (1982)). It is a well-known property of the Pareto optimum that

\[ \frac{de^3}{dY^3} = 1 \]  
\[ \text{the zero marginal tax at the top}. \]  

Adding the condition for Pareto optimality, we have five equations implicitly defining all other variables as functions of $V^3$ including $V^1(V^3)$, which can be interpreted as a Pareto frontier. Even if consumers of categories 1 and 2 have identical indifference maps and thus a utility function in $Y, B$-space with the same ordinal properties, the welfare weights that the government assigns to income changes for the two categories may be different and governed by some transformation of the ordinal

---

6 By substituting from (26) and (27) in (1) and equating the right hand sides of (27) and (28) we are left with two constraints and can maximise w.r.t. $Y^1$ and $Y^3$.

7 The intuition is the following. Type 1, being less productive, has less leisure for a given value of $Y$. Hence he requires a larger compensation in terms of B for a marginal increase in labour supply. Being less productive, he also has to increase working hours more in order to generate a unit increase in gross income. This reinforces the conclusion that a higher increase in B is required to offset the disutility from a marginal increase in $Y$, and the slope of the indifference curve must be larger.
utility function. We let the cardinalisation \( V^2 = \nu(V^1) \), with \( \nu' > 0 \), dictate the weight given to the high-skilled person with strong preference for leisure. By letting \( \nu' < 1 \), additional income for high-skilled with strong preference for leisure will be given less weight than additional income for low-skilled when the initial income point is the same.

We can now examine how the choice of tax policy will be affected by the composition of the low income group. Let \( m \) be the number of high-skilled individuals with low preference for leisure and let \( n \) be the number of low-skilled individuals, and \( M = n + m \). We can then examine how the composition of the \( M \)-group affects the tax design. By increasing \( m \) and lowering \( n \) correspondingly we see how taxes are affected by the fact that those with a low income consists not only of low-skilled persons, but also of high-skilled with a strong preference for leisure. The welfare function to be maximized is the sum of welfare across the three groups:

\[
\Omega = NV^3 + mV^2 + (M - m)V^1
\]  

(30)

From the analysis above we can substitute \( V^1(V^3) \) for \( V^1 \) and \( V^2 = \nu(V^1(V^3)) \) for \( V^2 \) and write the objective function as

\[
\Omega = NV^3 + m\nu(V^1(V^3)) + (M - m)V^1(V^3)
\]  

(31)

Maximising w.r.t. \( V^3 \) we get the first order condition

\[
\Omega' = N + m\nu'(dV^1/dV^3) + (M - m)\nu'(V^3) = 0
\]  

(32)

The second order condition is

\[
\Omega'' < 0.
\]  

(33)

By comparative statics we can examine the effect of an increase in \( m \) for a fixed \( M \) by differentiating the first order condition.

\[
\frac{\partial V^3}{\partial m} + (\nu' - 1)(dV^1/dV^3) = 0
\]  

(34)

It follows that

\[
\frac{\partial V^3}{\partial m} > 0 \text{ if } \nu' < 1.
\]  

(35)

The following proposition states the result.

---

8 A more general transformation might have been considered, but to avoid unnecessary complications we have chosen the simplest possible one for the purpose.
**Proposition 3:** If low-skilled agents and high-skilled agents with strong preference for leisure, have identical indifference curves in $Y,B$-space, and if high-skilled agents with a strong preference for leisure are given a lower welfare weight than low-skilled individuals, a larger share of high-skilled agents in the low-income group motivates shifting more of the tax burden to the low-income group.

In other words the tax should be less redistributive if more of the low-income people are in fact not low-skilled, but have a low income as a consequence of working few hours.

The problem of distinguishing the low-skilled and the high-skilled with strong preference for leisure does not necessarily vanish even if people with different preferences have deviating indifference curves in $Y,B$-space since it may not be optimal to separate the two types at an income tax optimum. It may be optimal to have bunching of the low-skilled and the high-skilled with strong preference for leisure even though their indifference curves through the shared income point in $Y,B$-space have different slopes. If information on preference type were available, it might be desirable to tax the high-skilled, leisure-prone type harder than the low-skilled, but in the absence of such information differentiated taxation may not be desirable.

The analysis above took as its point of departure the assumption of identical indifference curves of type 1 and type 2. However, what we actually used in the analysis was the fact that $Y^2 = Y^1$ and $B^2 = B^1$, or in other words a bunching assumption. The latter is a less robust assumption as in the absence of identical indifference curves there may be bunching at the tax optimum for some values of $\nu^3$ but not for others. We then have to take into account that in the course of the comparative static analysis the bunching may be retained or abandoned. However we can state.

**Proposition 4:** If high-skilled agents with a strong preference for leisure are given a lower welfare weight than low-skilled individuals, a larger share of high-skilled agents in the low-income group motivates shifting more of the tax burden to the low-income group as long as there is bunching of low-skilled and high-skilled with strong preference for leisure.
4. What can commodity taxes achieve?
It is well known from the tax theory assuming homogeneity of preferences that an important role for commodity taxes may be to alleviate self-selection constraints. (See for example Edwards et al. (1994)). This may be a role also in the case of heterogeneous preferences. A further role may be to provide a tool for taxing differently agents that the income tax is incapable of discriminating between. This may happen also in the conventional theory as there may be cases of bunching. However, we believe that this is a more pressing problem when preferences are heterogeneous as it is hard to distinguish between people who earn a low income because they are low-skilled and people who choose a low income because they prefer to work short hours. Unlike in the conventional theory different agents may even have preferences that make their indifference curves in $Y, B$ -space collapse into identical maps (as discussed by Boadway et al. (2002)). Already by inspecting the various cases of our taxonomy above we can say something about the potential role of commodity taxes.

Case i. In Case i market goods are separable from leisure in the utility function and preferences for commodities are the same. People who have the same disposable income will have the same commodity demands, and there is no way that one can differentiate commodity taxes between people who choose the same income point. There is no role for commodity taxes. In particular we observe that the only distinction between the low-skilled person and the high-skilled person with strong preference for leisure is that the latter, when selecting the same gross income, will have more leisure as, more productive, he will earn the same income in less time. However, the consumption bundles they choose will be the same as disposable income is the same, and, under separability of labour from market goods, the consumption bundle is unaffected by the amount of labour.

Case ii). In this case the utility function has the form $u'(f(z, h), h)$. In this case different types of people, who cannot be distinguished by observing gross incomes, will enjoy different consumption bundles when work effort differs, and hence they can be taxed differently by non-uniform commodity taxation. By taxing commodities that are complementary with leisure, a higher tax burden is imposed on those with a stronger preference for leisure. This case for commodity taxes resembles
the similar case under asymmetric information about skill level. Since a high-skilled person may like to mimic a low-skilled person with the same preferences by picking the same income point, there is an argument for taxing goods of which the potential mimicker has a high consumption as compared to the low-skilled person being mimicked. Those goods are exactly those that are demanded in larger quantities by people enjoying more leisure. In the present context high-skilled people with strong preference for leisure are motivated to mimic the low-skilled type by wishing to copy the choice of low income in order to satisfy their preference for leisure.

Case iii). If a person is endowed with preferences of the form \( u(f'(z), h) \), different preferences for leisure versus income will be due to the difference in preferences for market goods. Because of the separability of labour, differences in labour supply will induce no variation in commodity demands. However, even if people have the same income, they have heterogeneous preferences for market goods and will choose different consumption bundles. Since persons choosing the same \( B,Y \)-point may have different preferences for commodities, it follows that differential treatment can be achieved by means of commodity taxes.

Case iv) Both heterogeneity of labour supply and heterogeneity of preferences for market commodities imply that market commodity demands vary, and there is a potential role for commodity taxes in differentiating taxes.

4.1 Commodity taxes in the bunching case.

Let us now consider the role of commodity taxes when there is bunching of the low-skilled type and the high-skilled with strong preference for leisure. To simplify we confine the price setting to the choice of a commodity tax for just one good. We denote the quantity of this good consumed by a type \( i \) individual \( x^i \), the consumer price \( p \), and the tax \( t \). To maximise welfare subject to the budget constraint and self-selection constraint barring type 3 from mimicking, we formulate the Lagrange function:

\[
\Lambda = n^1 V^1 \left( P, Y^1, B^1 \right) + n^2 V^2 \left( P, Y^2, B^2 \right) + n^3 V^3 \left( P, Y^3, B^3 \right) - \beta \left( V^3 \left( P, Y^1, B^1 \right) - V^3 \left( P, Y^3, B^3 \right) \right) \\
+ \mu \left( n^1 \left( Y^1 - B^1 \right) + n^2 \left( Y^2 - B^2 \right) + n^3 \left( Y^3 - B^3 \right) + t \sum_{i=1}^{3} x^i \right)
\]

(36)
Since we are considering the bunching case we impose the conditions $B^1 = B^2$ and $Y^1 = Y^2$ and rewrite the expression above as:

$$
\Lambda = n'V^1 (P, Y^1, B^1) + n^2 V^2 (P, Y^1, B^1) + n^3 V^3 (P, Y^1, B^1) - \beta (V^3 (P, Y^1, B^1) - V^3 (P, Y^3, B^3)) + \mu \left( n^1 (Y^1 - B^1) + n^2 (Y^1 - B^1) + n^3 (Y^3 - B^3) + t \sum_{i=1}^3 x^i \right)
$$

(37)

where $P$ is the consumer price vector. The policy decision is to choose $B^1, Y^1, B^3, Y^3$ and $t$. The first order conditions with respect to $B^1, Y^1, B^3, Y^3$ are:

$$
\Lambda_{B^1} = n^1 V^1_B + n^2 V^2_B - \mu (n^1 + n^2) + \mu t x_1 n^1 + \mu t x_2 n^2 - \beta V^3_{B1} = 0
$$

(38)

$$
\Lambda_{Y^1} = n^1 V^1_Y + n^2 V^2_Y + \mu (n^1 + n^2) + \mu t x_1 n^1 + \mu t x_2 n^2 - \beta V^3_Y = 0
$$

(39)

$$
\Lambda_{B^3} = n^3 V^3_B - \mu n^3 + \mu t x_3 n^3 + \beta V^3_B = 0
$$

(40)

$$
\Lambda_{Y^3} = n^3 V^3_Y + \mu n^3 + \mu t x_3 n^3 + \beta V^3_Y = 0
$$

(41)

Using Roy’s theorem the first order condition with respect to $t$ ($p$) is:

$$
\Lambda_t = -n^1 V^1 x_1 - n^2 V^2 x_2 - n^3 V^3 x_3 + \mu \left( n^1 x_1 + n^2 x_2 + n^3 x_3 \right) + \mu t x \beta V^3_{1x1} - \beta V^3 x_3 = 0
$$

(42)

We will manipulate these first order conditions to obtain a formula for the commodity tax rate. When doing this the following expression will be useful

$$
\Lambda_{B^2} = n^2 (V^2_B - \mu (1-t x_2^2))
$$

(43)

This expression is obtained by differentiating eq. (36) partially w.r.t. $B^2$. By the envelope theorem this is the welfare effect of a hypothetical increase in $B^2$, i.e. the welfare effect that would be obtainable if it were possible to undertake a partial increase in $B^2$. It reflects the desirability, but not the feasibility, of increasing after
tax income (decreasing the income tax) for type two. Disregarding $n^2$, we have that the first part $V^2_b$ is the social benefit of more after tax income to a type two person, whereas the second part $-\mu(1-\mu^2_b)$ is the social cost, net of possible tax revenue effects of the commodity tax. In general we can not tell the sign of this. However, if we are in a situation such that we would like to tax a type two person harder and redistribute the proceeds to a low skill person, we would in general believe $\Lambda_{x^2}$ to be negative.

Using $s^i$ to denote the substitution effect for the taxed good for individual $i$, defining $S = n^1s^1 + n^2s^2 + n^3s^3$, using the first order conditions (38)-(42) and the Slutsky decomposition we can state the following proposition:

**Proposition 5:** Assume the income tax optimum is characterized by bunching of the low-skilled type and the high-skilled with strong preference for leisure, then we would like to tax a commodity according to the rule

$$t = \frac{\Lambda_{x^2}(x^2-x^1)}{\mu S} - \frac{BV^3_b(x^{31}-x^1)}{\mu S}$$  \hspace{1cm} (44)

**Proof:** See Appendix A

The second part of this formula we recognize as the usual “mimicking” term. The coefficient multiplying $(x^{31}-x^1)$ is positive. (This is because $V^3_b > 0$, $\beta$ and $\mu$ are positive lagrange multipliers and $S$ is negative.) Hence, the last term is positive if the “mimicker” consumes more of the good in question than the low skill person. We should use the commodity tax to deter mimicking, thereby mitigating the self-selection constraint. Since the mimicker has more leisure time than the low skill person the good should be taxed if the demand increases in leisure and subsidized if demand decreases in leisure. This is the same story as told in e.g. Edwards et. al (1994).

The first term, which we will denote the “redistribution” term, is of another character. The difference in consumption between the (actual) high skill person with a strong preference for leisure and the consumption of the low skill person is of
crucial importance. \( \Lambda_{g^z} \) shows how we value an increase in after tax income to a type two person. Let us for the moment assume this term is negative. Then the term \((\Lambda_{g^z} / \mu S)\) multiplying \((x^2 - x^3)\) is positive. If the type two person has a higher consumption than the low skill person, this term contributes to a higher value for \(t\). If the high-skilled type is the larger consumer one can impose a commodity tax and lower the income tax on the low-skilled to compensate him for the commodity tax burden. This tax relief will be inadequate to compensate the high-skilled type with strong preference for leisure when the latter is the larger consumer. Thus a partial, additional tax burden is imposed on the person, who was considered to have been taxed too leniently. Because of this we have chosen to denote this term the “redistribution” term.

To continue the discussion of formula (44) it is useful to use the taxonomy in section 2. If preferences are of the form in case \(i\), both the mimicking term and the redistribution term in (44) will vanish and nothing can be gained by a commodity tax. If the preferences are of the form in case \(ii\), the type two (actual) person and the (potential) mimicker will have the same consumption bundles. If \(\Lambda_{g^z}\) is negative the redistribution and mimicking terms will reinforce each other and they both imply that there should be a tax if demand increases in leisure and a subsidy if it decreases. In this sense the two terms are similar. However, note the difference that there will be no actual tax revenue from the mimicker, whereas the type two person actually pays the tax. Even though we regard it as the normal case that \(\Lambda_{g^z}\) is negative, we can not rule out the possibility that it is positive. In that case the two terms in eq. (44) would counteract each other.

If preferences are like in case \(iii\) in our taxonomy, i.e. preferences for type 3 and type 1 are of the form \(U(f(z), h)\), but for type two persons of the form \(U(\phi(z), h)\), then the mimicker term will vanish in eq. (44) but not the “redistribution” term. By exploiting the differences in consumption bundles one may impose commodity taxes on goods of which the high-skilled with strong preference for leisure have high consumption and possibly introduce subsidies favoring the low-skilled. Hence, even though we can not use the income tax to redistribute from the high skill person with strong preference for leisure and the low skill person, we can achieve redistribution via the commodity tax.
If the utility function is not separable and preferences are heterogenous, i.e. type 1 and type 3 have a utility function $U(z, h)$ and type 2 a utility function $\bar{U}(z, h)$, the mimicking term would point in the direction of taxing the commodity if the demand according to $U(z, h)$ increase with leisure. However, even if $\bar{U}(z, h)$ implies a demand function where $x^2$ is, say, increasing in leisure, it might still be the case that $x^2$ is less than $x^1$ and the redistribution term points to a subsidy.

5. Further on utility comparisons.
Decisions on distributional policy must involve some kind of interpersonal utility comparisons, even if solely comparing utility levels will in general not be sufficient for determining the preferred direction of transfers.

In order to compare the utility levels of different persons, say two persons labelled $a$ and $b$, we need to be able to determine how a consumption bundle $x$ for person $a$ compares to a consumption bundle $y$ for person $b$, or, provided that person $a$ gets a consumption bundle $x$, which is the consumption bundle $y$ which is equally good for $b$ as is $x$ for $a$? In general this is an extremely complicated task. To approach the problem it may be helpful to conceive of special cases in which this welfare comparison has a "natural solution." An example may be the case in which the two types of persons have symmetric preferences implying that if person $a$ consumes quantities $x_1, x_2$, for some given units of measurement, then person $b$ will be equally well off if his consumption bundle is $x_2, x_1$. (Of course, he will also be equally well off at other bundles on the indifference curve through $x_2, x_1$). If $a$ consumes 7 units of commodity 1 and 3 units of commodity 2, then $b$ will be equally well off if he consumes 3 units of commodity 1 and 7 units of commodity 2. (These may be assumed to be the only commodities, or consumption of other commodities is the same for both types.) Cases in which symmetry appears to be a plausible assumption are those in which the two goods are essentially variants of the same basic commodity, say, red and white wine. Person $a$ has a preference for red wine whilst person $b$ has a preference for white wine, but white wine is for person $b$ what red wine is for person $a$ and vice versa. Assuming symmetric preferences, the two types of consumers will obviously be equally well off if the the two commodities have the same price. If relative prices change they will no longer be equally well off.
In general the symmetry assumption is too restrictive, but also in general it appears a plausible assumption that for each indifference curve of a person labelled \( a \) there is an indifference curve of another person \( b \), which implies that when \( b \) is on that curve the two persons are considered to be equally well off. In particular there is a point of intersection, where both consume the same quantities of both commodities, such that the persons are equally well off. This is the reasoning underpinning the assumptions that have been made above.

We should note that in general the marginal utility of income does not only depend on the utility level, but also on the bundle of goods that generates the utility level. For a fixed utility level \( V \), and for convenience measuring labour supply by gross income \( Y \), we can express \( B \) by an expenditure function \( e(Y,V) \). The slope of the indifference curve in \( Y,B \)-space is \( dB/dY = e_Y \). Inserting \( V(B,Y) \) we can write \( B = e(Y,V(B,Y)) \). Differentiating partially with respect to \( B \) we find the marginal utility of income \( V_B = 1/e_Y \). To trace the variation along the indifference curve we consider

\[
\frac{\partial}{\partial Y} \frac{1}{e_Y} = -\frac{e_{V_Y}}{e_Y}.
\]

It is easy to recognise that \( e_{V_Y} > 0 \) if leisure is strictly non-inferior. The marginal utility of income will in general depend on the consumption-labour bundle, and it will decrease along an indifference curve as the labour supply and income increase if leisure is a strictly non-inferior good. It follows that a low marginal utility of income may be assigned to a person although he has a low utility level if he is sufficiently far out along his indifference curve.

The marginal utility of income will depend solely on the utility level (be constant along an indifference curve) only if there is no income effect on leisure \( (e_{V_Y} = 0) \). That means that the utility function belongs to the class \( V = \psi (B-g(Y)) \), and we see that \( V_B = \psi'(B-g(Y)) = \psi'(\psi^{-1}(V)) \).

An analogous approach may be used to discuss the marginal utility of leisure. Let \( Y = \varepsilon(B,V) \) be the labour (gross income) that in conjunction with a disposable income \( B \) will yield the utility level \( V \). \( \varepsilon_B \) is the inverse of the slope of the indifference curve in \( Y,B \)-space. Differentiation of \( Y = \varepsilon(B,V(B,Y)) \) implies

\[-V_Y = -1/\varepsilon_Y, \quad \text{and} \quad \frac{\partial}{\partial B} \frac{-1}{\varepsilon_Y} = \frac{\varepsilon_{BY}}{\varepsilon_Y^2}.\]

When \( B \) is a normal good \( \varepsilon_{BY} > 0 \), and the marginal
utility of leisure increases as one moves out along the indifference curve and $B$ increases.

We assume that the government is inequality averse in the sense that the social marginal utility of income is decreasing in utility for given ordinal preferences. We can write $V_b = V_b(B,Y) = V_b(B,\epsilon(B,V)) = \tilde{V}_b(B,V)$. By non-inferiority of leisure the marginal utility of income is non-increasing in $B$ for a fixed utility level. Also assuming that the marginal utility of income is decreasing in $V$, it follows that the marginal utility of income is decreasing in leisure, which increases $V$, and in $B$ both directly and because of the increase in $V$. This is the first part of Assumption AI. When preferences are $B$-quasi-linear, the marginal utility of income is uniquely determined by the utility level and the inequality aversion implies that the marginal utility of income is inversely related to the utility level as stated in the second part of Assumption AI. In other words, a lower marginal utility is always assigned to the person with a higher utility. It follows from the inverse relationship that a higher marginal utility always implies a lower utility level.

Our basic assumption is that a person with a particular preference is not necessarily better or worse off, and does not necessarily derive more or less utility from additional resources than a person with a different preference. The marginal utilities of additional resources depend on which indifference curves the respective persons are on, and also their location on the respective indifference curves. On top of this the form of the additional resources makes a difference as other things equal, or based on pure preferences, the individual with stronger preference for income would like to have additional income and the individual with stronger preference for leisure would like to have additional leisure. To make this notion of different tastes precise we need to isolate the effect of preferences, or in other words to define the meaning of ‘other things equal’.

A natural definition of equal starting points is that the two types have the same consumption bundle, and are equally well off. We assume that at this point of departure the type with stronger preference for leisure will acquire the larger increase in utility from a given increase in leisure, whilst the person with a stronger preference for leisure will obtain a larger increase in utility if they both acquire the same increase
in income. This is the rational for Assumption U above. When preferences are B-quasi-linear, the marginal utility of income is constant along an indifference curve. Then the type of individual with a stronger preference for income will always have the higher marginal utility of income as long as the persons remain equally well off. Thus Assumption UB follows as an implication of Assumption U and the B-quasi-linearity. However, as the marginal utility of leisure will change along the indifference curve we cannot tell in general who will have the higher marginal utility of leisure as we depart from the point of equal consumption bundles. A further implication for the case of B-quasi-linearity is that if the person with stronger preference for leisure in some situation has the higher marginal utility of income, he is also worse off. This is straightforward. If utility levels were the same he would have a lower marginal utility of income as cet. par. he derives less utility from income. Since the marginal utility of income is decreasing in utility, it is only if he has a sufficiently lower utility that he will have the higher marginal utility of income.

6. Conclusion.

We have highlighted the problems involved in tracing the implications of heterogeneous preferences for optimum taxes. In this context we have argued that it is helpful to classify various types of preference heterogeneity. Which case that prevails is shown to be crucial for what can be achieved by various tax instruments.

We have singled out two issues for further analysis. Firstly, we have characterised various income tax optima. In particular we have discussed the case in which there is bunching of low-skilled individuals and high-skilled people with strong preference for leisure, and demonstrated how the optimum income tax will then depend on the composition of the population. Secondly, we have discussed the role of commodity taxes. We have identified the preference cases in which tax differentiation not attainable by income taxes, can be achieved by means of commodity taxes. This may happen either because preferences for market commodities differ or because high-skilled individuals have more leisure and leisure in turn induce consumption of certain commodities (leisure goods). Finally, we have discussed the interaction of preferences for market commodities and the preference for leisure, which implies that the trade-off between income and leisure in general will be affected by prices and

---

9 Note that this does not follow from the ordinal properties, which only describe the relative valuation of the commodities.
hence can be manipulated by commodity taxes. Interestingly, this may provide a tool for separating individuals that would otherwise be bunched together at the income tax optimum, but only at a cost since commodity taxes are distortionary.

Appendix A

Proof of proposition 5.

Making use of the Slutsky decomposition we can rewrite eq. (42) as:

\[ \Lambda_p = -n^1V^1_B x^1_B - n^2V^2_B x^2_B - n^3V^3_B x^3_B + \mu \left( n^1x^1 + n^2x^2 + n^3x^3 \right) + \mu t(n^1s^1 + n^2s^2 + n^3s^3) \]

\[ -\mu \left( n^1x^1_B + n^2x^2_B + n^3x^3_B \right) + \beta V^3_B x^3_B - \beta V^3_B x^3 = 0 \]  

(A1)

where \( s^i \) is the Slutsky derivative of \( x^i \). Multiplying (38) by \( x^i \), we get

\[ n^1V^1_B x^1_B + n^2V^2_B x^2_B - \mu \left( n^1x^1_B + n^2x^2_B \right) + \mu t x^i(n^1x^1_B + n^2x^2_B) - \beta V^3_B x^3 = 0 \]  

(A2)

Multiplying (40) by \( -x^3 \), yields

\[ -n^3V^3_B x^3_B + \mu n^3x^3_B - \mu t x^3_B n^3x^3 - \beta V^3_B x^3 = 0 \]  

(A3)

We can make use of (A3) to eliminate a number of terms in (A1). Then adding each side of (A1) and (A2) we obtain

\[ n^3V^3_B (x^1_B - x^2_B) - \mu n^3(x^1_B - x^2_B) + \beta V^3_B x^3_B = 0 \]  

(A4)

where \( S = n^1s^1 + n^2s^2 + n^3s^3 \). Now returning to (36) and differentiating w.r.t. \( B^2 \), we find

\[ \Lambda_B^2 = n^2(V^2_B - \mu(1-tx^2_B)) \]  

(A5)

We can then reformulate (A4) as

\[ -\Lambda_B^2 (x^2_B - x^1_B) + \beta V^3_B (x^3_B - x^1_B) + \mu t S = 0 \]  

(A6)

which can be written as eq (44).

References


