IN PRAISE OF FRANK RAMSEY'S CONTRIBUTION TO THE THEORY OF TAXATION*

Joseph E. Stiglitz

Frank Ramsey's brilliant 1927 paper, modestly entitled, 'A contribution to the theory of taxation', is a landmark in the economics of public finance. Nearly a half century later, through the work of Diamond and Mirrlees (1971) and Mirrlees (1971), his paper can be thought of as launching the field of optimal taxation and revolutionising public finance.[1]

Ramsey, in his short life, made pathbreaking contributions to two other fields, the theory of optimal growth (Ramsey, 1928) and the theory of subjective probability (Ramsey, 1926).[2] Here, he addresses a question which he says was posed to him by A. C. Pigou: given that commodity taxes are distortionary, what is the best way of raising revenues, i.e. what is the set of taxes to raise a given revenue which maximises utility. The answer is now commonly referred to as Ramsey taxes. The basic insight was that taxes should be set so as to reduce the consumption of each good (along its compensated demand curve) equi-proportionately. He establishes this result in two contexts:

(*i*) if the government needs to raise only a small amount of tax revenue; and
(*ii*) if utility functions are quadratic.

The analysis is beautiful, mathematically sophisticated, making use of all the artillery in the theory of consumer behaviour, including the symmetry of the Slutsky relations. The conclusion overturned simplistic analyses that somehow continue to prevail in some quarters for decades after Ramsey's paper and, in some subfields of economics, continue to this day. Some economists argued against differential taxation on the grounds that it is best just to have a single tax (typically on wages). A wage tax introduces a single distortion – between the marginal rate of substitution between

[1] Given the huge number of papers that have been written in this field in the last half century, this article cannot even attempt a survey. Rather, I focus on the central themes raised by Ramsey in his paper. Textbook treatments are provided by Atkinson and Stiglitz (1980) and Salanié (2003).

[2] In addition, he made important contributions to mathematics, logic and philosophy, before dying at the young age of 26.

labour and consumption and the marginal rate of transformation. Interest income taxes and commodity taxes introduce additional distortions. (Interestingly, Ramsey's analysis has, wrongly, continued to be used by some economists to argue against the taxation of capital.)[3]

As another example of such simplistic economics, some macroeconomists continue to argue that monetary authorities should only interfere with the market in setting short-term interest rates.

Ramsey showed that efficient taxation required imposing a complete array of taxes – not just a single tax. A large number of small distortions, carefully constructed, is better than a single large distortion. And he showed precisely what these market interventions would look like. (He even explains that the optimal intervention might require subsidies – what he calls bounties – for some commodities. 'A tax on sugar might reduce the consumption of damsons more than in proportion to the reduction in the total consumption of sugar and so require to be offset by a bounty on damsons', Ramsey, 1927, p. 54). Similar reasoning would suggest that if there are a variety of tools for managing the macroeconomy, in general, they should all be employed.

In this short celebratory article, I briefly describe Ramsey's basic insights (Section 1), and the early history of the development of the ideas based on Ramsey's paper (Section 2). While several of these crucial developments showed that Ramsey's conclusions held under more general conditions than he had assumed, later analyses showed crucial qualifications, so that the policy relevance of Ramsey's analysis may be limited.

1. Major Insights

One of the striking aspects of the paper is that he not only provides a general formulation of the set of optimal taxes but he also translates this into concrete results in the case of linear demand and supply functions (quadratic utility functions) and shows how the results can be extended to situations when there are restrictions on the set of feasible taxes. In particular, when there are a set of commodities with fixed taxes (including commodities that cannot be taxed at all), he shows that there should be an equi-proportionate reduction in the goods for which taxes can be freely set. In the case of linear and separable demand and supply curves (quadratic utility functions) and small taxes, he shows that optimal taxes are inversely related to the compensated elasticity of demand and supply.

$$t/p = 1/\text{elasticity of demand} + 1/\text{elasticity of supply}. \tag{1}$$

This implies, of course, raising all the requisite revenue from a good which is either inelastically demanded or supplied, because such a tax generates no distortions. This in turn has one other implication: if say labour is supplied inelastically, then optimal taxation requires imposing a tax on labour alone; but a proportionate tax on all commodities is equivalent to a tax on labour. This is hardly a surprising result – one for which one did not need Ramsey's apparatus (as Ramsey himself recognised): when

---

[3] See the discussion below and the references cited there.

there is an inelastic factor, it has long been recognised that there is no distortion associated with taxing it.[4]

Ramsey's analysis can be viewed as the first successful exercise in second best economics. First best economics called for the imposition of lump sum taxes. Later, Meade (1955) and Lipsey and Lancaster (1956) would call attention to the fact that when some of the efficiency conditions are not satisfied, it may mean that others should not be satisfied. Their analysis would suggest that it would be very difficult indeed to characterise optimal resource allocations under such conditions. But Ramsey showed precisely how this could be done, when the reason for the distortion (the discrepancy between the marginal rate of substitution in preferences and the marginal rate of transformation in production) was to raise funds for the government.

Ramsey, however, went beyond this into an exploration of third best economics. He asked, what happens if there are some commodities that cannot be taxed, or whose tax rates are fixed. He argues that the same result (on the equi-proportionate reduction in consumption) holds for the set of goods that can be freely taxed.

Much of the huge subsequent literature on optimal taxation has been concerned with applying Ramsey's insights, for example, to the theory of capital taxation; extending the results, for example, to more explicitly general equilibrium models); and to understanding the limits of the results. This article briefly notes some key examples in each of these areas.

2. Earlier Dissemination

While Ramsey's work was incorporated into Pigou's (1928) classical public finance textbook, for several decades, his work had limited impacts on the broader, and more institutional, field of public finance. Interestingly, one of the early developments in the design of optimal taxes (by Corlett and Hague, 1953) was a special application of Ramsey's analysis, without fully realising it. They argued that if leisure cannot be taxed, then one should impose surrogate taxes, taxing complements to leisure at a higher rate (and conversely, substitutes should be taxed at a lower rate). The extreme case is easy to see: if the consumption of some good was perfectly associated with leisure, then a tax on that good is fully equivalent to a tax on leisure.

Ramsey's results were seemingly independently discovered by Boiteux (1956),[5] in the context of regulatory policy – analysing the optimal prices for a regulated utility which has to raise revenues to pay the fixed costs of the utility. The basic insight was fully parallel: prices (relative to marginal costs) should be raised so that there was an equi-proportionate reduction in consumption along the compensated demand curve.

It was a short distance from there to the observation that a multi-product monopolist would charge Ramsey-like prices. The fact that a monopolist and a welfare maximising government would impose similar prices suggested that perhaps monopoly was not as

---

[4] Henry (1886) focused on the taxation of land as the inelastic supply factor of production. Subsequent research by Vickrey (1977, 1999) and Arnott and Stiglitz (1979) identified conditions under which such a tax would provide all the revenue that was required for financing public goods (in economies in which there was a local public good and congestion or other factors leading to decreasing returns to scale led to optimal sized communities.)

[5] English translation 1971. See also Baumol and Bradford (1970).

bad as had previously been thought; and that conclusion became reinforced in those contexts where monopoly profits are driven to zero by contestability[6] (i.e. as firms compete for the market) – or in contexts where the government auctions off the right to be the monopolist; for then it would appear that the monopoly pricing is identical to that which would arise in the case of a 'social planner' maximising the utility of the representative agent, subject to the constraint that monopoly profits are zero, i.e. no lump subsidy to the firm is allowed to offset its fixed cost. I say 'appear', because the conclusion is not in general true, as Sappington and Stiglitz (1987a,b) explain. For instance, the government would take into account the effect of any change in the prices it charges on the revenues that the government raises from taxes or other sources; the monopolist would not. The government would be sensitive to the distributional consequences of alternative pricing structures; the monopolist would not.

3. Diamond–Mirrlees

Ramsey's influence on modern public finance has been largely mediated through the foundational paper of Diamond and Mirrlees (1971) (DM), which set Ramsey's problem in the context of a fully specified general equilibrium problem. Assuming a standard constant returns to scale neoclassical production function and that linear taxes could be levied on every commodity (but that lump sum taxes could not be imposed), they showed that Ramsey's results held at the margin even when the revenues raised by government were large. (There were many other insights to their analysis – most importantly, that one only wanted to impose taxes on consumption; one did not want to impose any taxes that would interfere with production efficiency.)[7]

Though the most general Diamond–Mirrlees's result on equi-proportionate reductions in output of each commodity along the compensated demand curves[8] seemed fully parallel to that of Ramsey, the exploration of specialised applications suggested what might at first seem a puzzling difference: with separable demand functions, while Diamond–Mirrlees suggested that tax rates ought to be inversely proportional to the elasticity of demand (of the compensated demand curve), Ramsey had suggested that it was related both to the elasticity of demand and supply. Ramsey's result seems intuitive: there is no distortion associated with taxing a good with a zero supply elasticity. The two results are reconciled, once one takes into account the particular technology employed by Diamond–Mirrlees. In DM, there are no rents. In effect, supply curves are infinitely elastic.

4. Profits and Restricted Taxation

That raised a question: what is the optimal structure of commodity taxes if there are profits. Ramsey had posed another question: what do third best tax structures look like,

---

[6] See, e.g. Baumol et al. (1982). For a survey, see Brock (1983). As Farrell (1986) and Stiglitz (1988) point out, markets will not be contestable if there are even arbitrarily small sunk costs.
[7] As we note below, this result was not general.
[8] As Dasgupta and Stiglitz (1971) point out, what is required is an equi-proportionate reduction in consumption relative to what it would have been if consumer prices were at the after-tax producer prices. The after-tax producer prices may differ markedly from the before-tax producer prices.

when one cannot impose lump sum taxes but one is also restricted with respect to the set of (linear) taxes that one can impose on other commodities? Ramsey had shown that his equi-proportionate result still holds for taxes on those commodities which could be freely taxed. Was this true in a general equilibrium model? More generally, the question was posed: what do various restrictions on the set of taxes that can be imposed imply for the structure of commodity taxes and: what are plausible restrictions on the set of taxes? Identifying the feasible set of taxes was key to the analysis of the optimal tax structure. For if lump sum taxes were feasible (which would naturally be the case if all individuals were identical, as Ramsey had hypothesised), then there would be no need to resort to distortionary commodity taxes.

Dasgupta and Stiglitz (DS), who had been working in Cambridge as Diamond and Mirrlees were engaged in their pioneering work, were able to provide answers to these questions and, in doing so, qualified both the work of Ramsey and DM. When there is diminishing returns to scale (or there is a fixed factor) there will be rents (pure profits) and, as we noted earlier, when there are rents, they should be taxed first, before one turns to distortionary taxes. DS suggest three other general principles, parallel to the insight of Corlett and Hague (1953):

(*i*) use factor or commodity taxes to offset the effects of factors or taxes that cannot be taxed, even if doing so results in production inefficiency or abrogates the 'equi-proportionate reduction in consumption' rule;

(*ii*) especially increase taxes on products that provide an indirect tax on rents/profits; and

(*iii*) but to the extent that there are taxes on profits/rents, take into account the effect of the commodity tax on the revenues raised from the profits/rent tax.[9]

DM's analysis had been extremely influential in arguing against some important categories of taxes that interfere with productive efficiency. Tariffs on produced goods distorted production towards home-produced goods. The corporation tax seemingly distorted production away from goods produced in corporations. By contrast, DS showed that third (or fourth) best taxation may require taxation that interferes with production efficiency.[10]

Some examples illustrate: consider petrol. Petrol used by consumers can be considered an input into the production of transport services – that is the final consumption good being produced. Production efficiency requires not taxing an input, but taxing the output (transport services) but that output cannot be taxed (i.e. one cannot observe directly the amount the car is driven).[11] Hence, it is 'optimal' to tax the input of petrol as a surrogate for the (missing) tax on the output, even though

[9] This result was very much in the spirit of Ramsey and DM, who emphasised the effect of an increase in the tax on one commodity on the revenues raised by other taxes.
[10] At the same time, they showed that when the rents of different producers can be taxed at different rates, production efficiency is still desirable.
[11] Even if one could observe the total amount driven, that would not be a good measure of the 'output'. A taxi cab charges different amounts for driving at different times of the day and the fees reflect too the degree of congestion, so too, in principle, should we want to differentiate the amount driven in different circumstances. These are different 'outputs', and if we do not tax different outputs at different rates, we are imposing a constraint on the set of feasible taxes. DM assumed no such constraints; the discussion below describes the implications.

doing so introduces a production inefficiency. If we cannot tax the pure profits (rents) arising in the oil industry, it may be optimal to impose a (higher) tax (than we otherwise would) on the output of the oil industry (petrol). On the other hand, if we do tax the pure profits (rents) in the oil industry, we need to adjust the optimal tax on petrol to take into account the lower revenues that will result as a consequence of the tax.

These results are of particular relevance in developing countries, where it may not be possible to tax the output in agriculture, especially goods that are sold within the rural sector. But it may be possible to tax some of the inputs. It may be possible to tax, for instance, fertiliser, or imports of tractors. Doing so would clearly cause a production distortion. But as a third best policy, it may still be desirable to do so (Emran and Stiglitz, 2005; Stiglitz, 2010. For an example of the controversy to which this gave rise, see Keen, 2008).

In some special cases, one can obtain results of a remarkably simple form analogous to those obtained by Ramsey. DS showed that with a pure profits tax at the rate τ, in the special case of his independent demand curves, we obtain instead of (1):[12]

$$t/p = 1/\text{elasticity of demand} + (1 - \tau)/\text{elasticity of supply}. \qquad (2)$$

With a 100% tax on pure profits or infinite supply elasticities, we get a result analogous to DM, where it is only properties of the demand function which affect tax rate. With a zero tax on pure profits, we get the original Ramsey result. Ramsey had implicitly assumed that all profits were distributed to the owners of the firm, and not collected by the government. For finite supply elasticities, the higher the rent tax, the lower the commodity tax rate, because of the adverse effect of the increase in the commodity tax on the government revenues from the rent tax.

One might well ask, why not impose a 100% tax on profits? The answer brings us to what I have emphasised is the key issue in determining the set of feasible taxes – itself central to the question of the design of the tax system: imperfect information and what is observable. Earlier, we noted that in developing countries, taxation in the rural sector is limited: it is not possible for government to observe most transactions, for example, among farmers (Sah and Stiglitz, 1992). By the same token, we know that some of the returns to a firm in excess of payments to labour and other inputs is a result of entrepreneurial effort; it is not just pure profits, or even monopoly profits. While the government cannot identify what part is due to entrepreneurial effort, it knows that a 100% 'profits' tax would squelch effort. The optimal profit tax rate takes into account the impact of that tax on effort (Mirrlees, 1972).

DS's analysis of the analysis of commodity taxation with imperfect profits tax seemed much out of the spirit of Ramsey's important analysis of the design of commodity taxation when there were some commodities that were not taxed (or whose tax rates were fixed). He suggested that this additional restriction did not affect the optimal taxation of those that could be taxed. DS showed that his conclusion was not general: it only held if the demand for the taxed goods does not depend on the prices of the untaxed or partially taxed goods and *vice versa*. The intuition follows from what we have said above: one of the goods for which taxes are free to adjust may be a good surrogate

---

[12] These are compensated elasticities.

for the untaxed good, and it may be desirable to use this surrogate. Alternatively, an increase in taxes on the freely taxed goods may lead to a reduction in revenues on the goods with fixed taxes, and clearly the government would take that into account.

There is another set of restrictions that Ramsey did not consider but is important in practice: there may be a group of commodities that may have to be taxed at the same rate, even though their demand (and where relevant, supply) properties may differ, and, accordingly in the Ramsey analysis, if it were possible to tax them at different rates, these goods should be taxed differently.[13] DS provide a simple formula for how that tax should be set: one forms a composite commodity (weighing the different goods by the consumer prices in equilibrium) and the reduction in consumption of this composite commodity should be the same as it is for other commodities.

But once again, the general principle of restricted taxation applies: if it is possible to indirectly tax some product within the composite which in an unrestrained world would have faced a higher tax, it is desirable to do so, for example, by taxing factor inputs into that sector – even if doing so leads to production inefficiency.

5. Distributional Considerations

There was something distinctly unpleasant about Ramsey's recommendations: it suggested taxing the necessities of life, which tend to have low price elasticities (especially for the poor) at a high rate. If such a policy were pursued, it would imply that taxation should be regressive.

Of course, Ramsey considered a model in which all individuals were identical, so distributional issues simply did not arise. But if all individuals were identical, there would be no problem imposing a lump sum tax. The only reason that we resort to distortionary commodity taxes is implicitly because we are concerned with the regressivity of a lump sum (*per capita*) tax. But that in turn means that we have to focus explicitly on the distributional consequences of commodity taxation.

When we do this, not surprisingly we obtain markedly different results. There is a tension between imposing high taxes on commodities with low (compensated) price elasticities (generating low dead weight loss) and imposing high taxes on commodities with high income elasticities – resulting in a more progressive tax system. (The precise formulas are given, for example, by Atkinson and Stiglitz (1972, 1980), and depend on the weight given to equity.)[14]

But if we are concerned with equity, in general, there is no reason to limit ourselves to commodity taxation. Ramsey's analysis can, as we observed earlier, be viewed as the first (and one of the best) examples of second best welfare economics. He assumed that lump sum taxes were not feasible and asked what is the best set of distortionary commodity taxes. But in the theory of the second best, it is always necessary to be precise about what are the constraints: even if lump sum taxes are not feasible, other

---

[13] There are a couple of reasons for this restriction. Partly it is administrative: the cost of having millions of tax rates, one for each precisely identified commodity, would be large. Partly it is based on information: private parties would have an incentive to try to get their products classified as one of the lower taxed products. It is costly for the government to gather the information required to implement and enforce fine differentiations.

[14] See also Diamond (1975) and Mirrlees (1975).

taxes may be. Most importantly, the government can impose redistributive income taxes, or even possibly redistributive wage taxes.

Going beyond the analysis of taxation with identical individuals, one faces a problem: how does one weigh the well-being of different individuals? The standard approach in the theory of optimal taxation developing out of Ramsey's work was to use a utilitarian social welfare function, i.e. simply 'add up' the utility of different individuals. But it turns out that many, if not most, of the qualitative results can be extended to Pareto efficient tax structures, tax structures such that no one (no group) can be made better off without making someone else – or some group – worse off.

Thus, the set of Pareto efficient tax structures depends explicitly on the set of feasible taxes.[15] If the number of hours worked is observable, then it is possible to have progressive wage taxes. If only income is observable (the product of wages per hour and the number of hours worked), then it is possible to have a progressive income tax but not a progressive wage tax.

But in the presence of a progressive income tax, the optimal set of commodity taxes looks markedly different from that suggested by Ramsey. In one polar case, where (the entire vector of) consumption is separable from leisure (work), Pareto efficient taxation implies no commodity taxation (Atkinson and Stiglitz, 1976; Mirrlees, 1976).

Kaplow (2006) shows that under the separability condition, whenever there is distortionary commodity taxation, a distribution neutral Pareto improvement can be obtained through the income tax system. Without separability or in the presence of consumer heterogeneity, however, commodity taxation may be desirable: if richer individuals consume proportionately more chocolate, or if chocolate is a substitute for leisure, a chocolate tax may enable one to obtain a desired redistribution with less distortion in the income tax system.

Though the case of separability is very special, what the Atkinson–Stiglitz theorem illustrates is that in the presence of an (optimal) income tax, the optimal set of commodity taxes looks markedly different from that described by Ramsey.

In particular, commodity taxation can be viewed as a particular type of Pigouvian corrective tax.[16] The focus is not on the impact on tax revenues, or even directly on dead weight losses (as usually conceived), but on impacts on the self-selection constraints that are central to the design of the optimal income tax. 'Loosening' the self-selection constraints has a first order effect on welfare, while the distortions associated with small commodity taxation have a second order effect on welfare.[17]

6. Non-linear Taxation and Heterogeneous Individuals

Moreover, once one admits the possibility of non-linear income taxes, should not one consider the possibility of non-linear commodity taxes? If so, the problem that Ramsey solved was a special case of a more general problem (just as the optimal linear income

---

[15] The concept of Pareto efficient taxation is noted in Mirrlees (1976) and Stiglitz (1982a), and explored further in Stiglitz (1985, 1987, 1998) and Brito et al. (1990). See also Konishi (1995) and Saez (2002).

[16] For a more general discussion of Pigouvian corrective taxes, see the commentary below.

[17] This is the essential point of Greenwald and Stiglitz (1986) and Arnott et al. (1994).

tax (Stiglitz, 2009) is a special case of the optimal non-linear income tax (Mirrlees, 1971)). Of course, as we have already noted, the special case may be the more relevant case. For many commodities, the possibility of arbitrage means that taxes have to be linear; but there are some commodities, perhaps most importantly, electricity, for which arbitrage opportunities are typically limited.

Most of the work in the analysis of optimal taxation with heterogeneous individuals assumes that individuals differ in simple ways, for example, their productivity, but not their preferences. This enables us to elide the classical welfare question of interpersonal utility comparisons. There is a simple question that has to be answered for determining the optimal degree of progressivity: if (full) income increases by 1%, by what percentage does the marginal social utility of income decrease? A broad consensus has emerged within the fraternity of optimal tax economists that the answer (related to revealed individual behaviour, e.g. in the context of risk taking) is by 1–2%. But if individuals have different indifference curves (curves that cross, perhaps even more than once), there is no such easy trick available. Changes in relative prices can have markedly different effects on different individuals, even with roughly similar income. In practice, we give considerable weight to the current set of consumer prices (affected as they may be by government policy) and assess the welfare impact by looking at the income equivalent changes arising from the proposed change in tax (or expenditure) policy.[18]

7. An Information Theoretic Perspective

There are many easily observable ways in which individuals differ that are easily observable, for example, age, gender, marital status, some of which (like age and gender) are not (easily) alterable. In principle, if the unobservable or imperfectly observable characteristics that we would like to base taxation on are correlated with observable characteristics or behaviour, there is some presumption that we should differentiate taxes based on those observables – unless all the relevant information is contained in some other variable that might more easily be the basis of taxation.[19] Thus, with separability of utility (between consumption and leisure) and the only basis of individual differences being labour productivity, an optimal progressive income tax fully captures the relevant information. But if more productive individuals (and only more productive individuals) eat chocolate ice cream rather than vanilla ice cream, then it might be desirable to supplement the optimal income tax with a tax on chocolate ice cream.[20] Similarly, we might feel that individuals have relatively little choice about how much they spend on heating – the main determinant is the weather – but it is not possible to observe the weather conditions

---

[18] Implicitly, the standard approach sets utility levels of two individuals with different preferences but the same income (at current wages and prices) as the same. The point is that we might come to very different views of the distributional impacts if we set the utility levels to be the same at different wages and prices, and assess the income equivalent impact of a change in a tax off of that different base.

[19] It is an open question why more extensive use of such variables is not made. In practice, some redistribution for the elderly is done through the social insurance system.

[20] For an early formal discussion of this point, see Mirrlees (1976); for a more recent discussion, see Saez (2002).

under which each individual lives. But fuel consumption may be (imperfectly) correlated with the weather, especially for poor individuals. Of course, some individuals may consume more fuel because they like a warmer temperature, and it is inefficient to subsidise fuel consumption, because it leads to excessive consumption. Still, an optimal tax structure might include a fuel subsidy, at least for poor individuals. While economists have typically criticised these subsidies, the optimal tax theory growing out of Ramsey's seminal work provides a rationale.[21]

8. General Equilibrium Incidence

We have described how at the centre of modern optimal tax theory and the work growing out of Ramsey lies a balancing of distributional and efficiency concerns. The early literature essentially took the before tax-and-transfer distribution of income, or more accurately, the relative wages of different individuals, as given. It worried about how an increase in taxes on the rich might affect their labour supply and thus the supply of funds available for redistribution. But government tax and expenditure policies can change the before tax and transfer distribution and, in particular, relative wages. This can, reduce the burden imposed on distortionary redistribution policies. If it is possible to do so, governments should – again, even possibly if it doing so interferes with production efficiency, or seemingly reduces the progressivity of the tax system.

Thus, assume there are two kinds of labour, skilled and unskilled, which are complements. By encouraging greater supply of skilled labour, the relative wages of the unskilled labour will increase. If skilled labourers have by and large higher incomes, this suggests that they be confronted with lower marginal tax rates but, if possible, with higher average tax rates, than they otherwise would have faced.[22]

9. Interest Income Tax

Ramsey's analysis also had important and profound implications concerning the taxation of interest income, which he himself recognised. We can treat current and future consumption as two different commodities.[23] Taxing interest income can (at least in some simple models) be viewed simply as a tax on future consumption. The question, then, is do they have the same price elasticities? If not, then they should be

---

[21] Much of the criticism though is based on political economy concerns that go well beyond the simple normative framework posited by Ramsey. The magnitude of the subsidies, once enacted, often go well beyond anything that could be justified by an optimal tax framework and are hard to reverse when circumstances change.

[22] See Stiglitz (1985, 1998). An increase in the average tax rate and a decrease in the marginal tax rate will ensure that income and substitution effects reinforce each other to increase the supply of skilled labour. In some simple models where, in the absence of general equilibrium effects, the limiting marginal tax rate is zero, in their presence, it may be negative. There is, of course, a large body of work on general equilibrium incidence, going back to the work of Harberger (1962) and of Mieszkowski (1969). In the context of the subject at hand, see Diamond (1978) and Naito (1999).

[23] Ramsey treated savings and consumption as two different commodities and argued that capital should be taxed but at a lower rate than wage income. Later literature paid relatively little attention to Ramsey's formulation.

taxed at different rates. (*A priori*, it is possible that future consumption be taxed at a lower rate, i.e. that there be an interest income subsidy.)

In the standard model with infinite lived individuals with constant utility discount rate, the long-run supply elasticity of capital is infinite, which means that there should be no taxation of capital.[24]

The Atkinson–Stiglitz result that with separability and an optimal (wage) income tax, there should be no commodity taxation had a deeply disturbing corollary: there should be no taxes on interest income. A host of conservative economists have exploited these results, without fully understanding their limitations, to inveigh against capital taxation, including inheritance taxation.[25]

These results, though, need to be qualified, for several reasons, related to the general observations made earlier. We might, for instance, like to tax wealthier individuals at a higher rate, particularly if they are in part wealthier because they have received large amounts of human or financial capital from their parents. Levelling the playing field – providing more equality of opportunity – necessitates a tax on such cross generational distributions. We cannot, however, monitor such transfers perfectly but capital income may be correlated with such transfers (especially at the upper end) and, if so, they provide an appropriate basis for taxation.[26]

Piketty and Saez (2012) also explain how if there is uninsurable uncertainty about future returns, capital taxation can be viewed as providing insurance against returns (similar to a point made by Domar and Musgrave, 1944).[27] Aiyagari (1995) and Chamley (2001) shows that the result does not hold when insurance markets are incomplete and their borrowing constraints. Ganesa *et al.* (2009) obtain corresponding results in an overlapping generations model.

---

[24] Judd (1985, 1999), Stiglitz (1985) and Chamley (1986). As Piketty and Saez (2012) point out, there are several reasons for this result, including the particular formulation of the welfare criterion. Stiglitz (1985, 1998) shows that the results hold even in an overlapping generations model.

[25] See, for instance, Lucas (1990), Atkeson *et al.* (1999) and Mankiw *et al.* (2009).

[26] A variety of papers have explored these ideas. See in particular Cremer *et al.* (2003), Farhi and Werning (2010) and Piketty and Saez (2012). As Piketty and Saez (2012) put it: 'In sum, two-dimensional inequality requires two-dimensional tax policy tools'. Piketty and Saez assume that the reason some individuals give more bequests than others is that they have a 'taste' for bequests. Earlier literature on bequests argued, however, that one of the main reasons for differences in bequests related to uncertainty about the timing of death and the absence of good annuities, in which case, adverse effects of taxing bequests will be limited (Stiglitz, 1978). A key question in analysing bequest behaviour is why such a large fraction of the population leaves a near-zero bequest and a key challenge is to reconcile the predicted wealth distribution with the observed wealth distribution, i.e. how to reconcile observed patterns of income and wealth distribution. In a model with i.i.d. distributions, with even a modest amount of caring for one's descendants one would expect a larger fraction of the population providing bequests than is in fact observed. Bevan and Stiglitz (1979) assume that the stochastic process for abilities exhibits serial correlation but regression towards the mean, but that negative bequests are not feasible; and in such a model, they are able to explain some of the key observed features. They did not, however, explore optimal tax implications.

[27] The implications for tax policy were pursued in several subsequent papers, e.g. Stiglitz (1969, 1972). After noting this as a potential role for taxation, Stiglitz (1985) raises a question about the reason that such risks are uninsurable. If it is because of potential adverse incentive effects, then in effect government provided insurance may be little better than private insurance, i.e. there may be the same adverse effects. Government provided risk sharing may simply replace privately provided risk sharing. It should be noted that the extent to which capital taxation performs this risk-sharing role depends on details of the tax code, e.g. the extent to which there are loss-carry forward and back provisions.

General equilibrium effects provide a more ambiguous basis of capital taxation. If capital is a substitute for unskilled labour (as seems increasingly the case with robotisation) but a complement of skilled labour (an implication of skill-biased innovation) then a capital tax might not only improve the after-tax distribution of income but also the before tax distribution of income. There can also be distributional effects through the demand side, for example, taxation of capital intensive consumption goods and subsidising unskilled labour intensive goods may lead to an improvement in the before-tax-and-transfer distribution of income.[28]

One of the most important reasons for taxing capital income is that we cannot clearly distinguish capital income from wage income, particularly the labour that goes into managing capital. When an investor gets an above average return, should the difference be viewed as a return to his skill as an investment manager and, therefore, really be viewed as a return to labour? (This question has significant policy resonance: in the US, private equity managers have attempted to receive capital gains treatment for their income, which in most cases is simply a return to their labour. More generally, when we tax capital income preferentially, there are efforts to convert wage income into capital income.)[29,30]

There is a still further reason for imposing taxes on capital, at rates equal to or greater than the rates for individuals at comparable income derived from labour. Ramsey, and most of the subsequent literature, assumed that markets were competitive, that the economy was well-described by a neoclassical model. But today, it is increasingly being recognised that much of the income at the top is derived from rents (monopoly rents, from the ability of corporate CEOs to appropriate for themselves significant fractions of corporate income, from political connections that allow some to get overpaid for selling goods and services to government or to underpay for the acquisition of assets from the government etc.). These rents are capitalised and the returns are treated as if they were returns to capital. While it may not be possible to identify precisely what income is 'rent', the taxation of the capital income of high-income individuals at high rates may actually discourage such rent-seeking activity and increase both equality and economic efficiency.[31]

---

[28] Stiglitz (1985, 1998) emphasises, however, that what matters (in the simple models explored in that paper) is how the tax on interest income affects the self-selection constraint. In an overlapping generations model, government may use either social security or debt policy to control capital accumulation. The paper also explores the case where the government can only affect savings through tax policy, in which case the direct effects described above become relevant.

[29] There are several other arguments for the imposition of taxes (or subsidies) on capital income and/or bequests. Piketty and Saez (2012) discuss several papers focusing on the effect on capital constraints or on the risk individuals might face in the labour market. With dynastic utility functions, bequests smooth consumption across generations, and thus increase average expected utility. If this were the only effect, depending on the welfare criterion hypothesised, it might even be desirable to subsidise bequests. For a general survey of inheritance taxation from an optimal tax perspective, see Cremer and Pestieau (2006). After all, bequests generate utility not only to the giver but also to the receiver. Moreover, one of the reasons for bequests in a dynastic model is to share one's good fortune with future generations. Restricting intergenerational consumption smoothing through bequest taxation thus results in less sharing, i.e. more volatility in consumption (Stiglitz, 1976; Bevan and Stiglitz, 1979; Farhi and Werning, 2010).

[30] Still another reason for imposing capital taxation arises when the social rate of time preference is less than private individual's, which can easily arise in overlapping generation models. See De Bonis and Spataro (2005). This distinction has played an important role in the theory of discounting for social cost–benefit analysis. See, e.g. Stiglitz (1982b).

[31] For a general theoretical discussion of these points, see Stiglitz (2012). For empirical evidence in support, see Piketty et al. (2011) and Piketty (2014).

10. Ramsey Taxation and Pigouvian Corrective Taxation

Ramsey essentially assumed a well-functioning market economy.[32] The only distortions in his model were the tax distortions upon which he focused. His mentor, A. C. Pigou, had highlighted the role of corrective taxation – using taxes to equal private and social costs, e.g. in the case of environmental externalities.[33]

Dasgupta and Stiglitz's (1971) analysis provided a hint of what optimal taxation might look like in the presence of other distortions. They showed that in their model, optimal taxes required an equi-proportionate reduction in consumption from what it would be if consumers were charged the (appropriately calculated) shadow prices.

Sandmo (1975) extended the analysis of Ramsey and the subsequent optimal tax literature including that of DS to include externalities, deriving Ramsey-like formulae which explicitly pay attention to the effects of corrective taxation on revenues derived from other taxes and to the distributional impact of taxation and of the externalities themselves (which may affect those at different levels of income differently). Sandmo (1976) showed, analogous to the results of DS, that if there are restrictions on the set of taxes that can be imposed, it will in general be desirable to impose taxes or subsidies on related goods, which discourage the consumption of the externality-generating commodity.

Cremer *et al.* (1998) then extended Sandmo's work to include the further insights that the later literature had provided, i.e. the dependence of optimal structure on the set of feasible taxes. They showed, for instance, that optimal taxation entails a general income tax combined with standard Pigouvian taxes under the Atkinson–Stiglitz assumptions of separability.[34]

Arnott and Stiglitz (1986) obtained formulae describing optimal taxes in the special case of distortions arising from moral hazard.[35] Insurance induces individuals to undertake too little effort at accident avoidance (increasing, as a result, the probability of the insured against event). They establish the (not surprising) result that taxes and subsidies should be used to discourage, for instance, smoking and drinking – consistent with the popular 'sin taxes', but without the moral overtones. But for our purposes, perhaps the most interesting insight arises in the context of what they call the three-level agency problem: governments attempt to affect both the behaviour of consumers directly and of insurance companies; and insurance companies attempt to affect the behaviour of those they insure. As they observe (p. 9):

> the determinants of optimal taxes in the three-tier problem are 'an order more complicated' than in the two-tier problem, for example depending on the rate of change of elasticities in addition to elasticities.

[32] Though in one of the examples he uses to illustrate the generality of his model, he does discuss the need to impose motor taxes to 'offset' damage to roads, and that his taxes should be viewed as added on to these corrective taxes.

[33] This Section focuses on the use of taxes to correct negative externalities. But subsidies can similarly be used to correct positive externalities, such as those associated with innovation. Recently, Stiglitz (2014) had derived Ramsey-like formulae for such corrective taxation.

[34] That is, if individuals of different types have the same marginal rates of substitution at any given consumption bundle.

[35] Many other special cases have also been examined. For instance, Mayeres and Proost (1997) analyse optimal taxation with congestion type externalities.

There has been considerable controversy over whether and how various results in optimal tax theory apply in the presence of externalities and other market failures. DS had earlier shown that in the presence of market distortions, the standard results in optimal tax theory should be viewed as describing deviations between consumer prices and *shadow* producer prices. Kaplow (2012) has similarly demonstrated analogous results in the presence of an income tax. One can think of designing the tax system as a two stage process–first imposing corrective taxes, to make producer prices aligned with shadow prices, and then going from these prices to consumer prices in the usual way. This means that with separability between leisure and consumption goods (including pollution), and an optimal income tax, no *further* commodity taxes should be imposed in the second stage; and if there are commodity taxes, there is always a distributional neutral reform of the income tax system that is Pareto improving. As before, it might be desirable to impose *additional* taxes (beyond the Pigouvian levels) on a pollution generating activity in the absence of separability or with preference heterogeneity.[36]

11. Ramsey Taxation and Developing Countries

Our discussion so far has suggested that Ramsey's results have limited direct applicability to the design of commodity tax structures (or to whether it is desirable to tax interest income) in advanced countries, simply because in almost all there are (or could and should be) progressive income taxes. But such taxes play a far less important role in most developing countries and it is in this context that they may be most relevant.

A careful application of Ramsey's analysis to these countries requires, however, a careful analysis of:

  (*i*) the constraints on taxation;
  (*ii*) the elasticities of demand and supply; and
  (*iii*) the structure of the economy.

As we have repeatedly noted, Ramsey's original analysis included a discussion of the design of taxation, given that certain taxes were fixed (and more relevant for our purposes, certain sectors/commodities were not taxed at all, i.e. face a fixed tax rate of zero). In developing countries, output and consumption in the rural sector typically cannot be taxed and often the only taxes that can be imposed are those on imports. Taxing imports, though, leads to a production distortion. Partly on the basis of these

---

[36] In particular Stiglitz (2013) constructs a model without separability showing that so long as the wage elasticity is positive and the revenues of the incremental pollution tax are used to reduce the income tax, there is a 'double dividend' from taxing pollution at the margin. (Nordhaus, 1993 has similarly argued that the revenue benefits of a carbon tax lead to a higher optimal level of the carbon tax). As we noted earlier, in the general case, the imposition of any tax has to be sensitive to the impact of the tax on other sources of revenue and about the distributional consequences of the tax and the uses to which any revenue raised is put. What it neat about Kaplow's formulation focusing on distributional neutral changes is that he addresses these issues head on. It appears that those, like Parry (1994) and Bovenberg and Goulder (1996), who argue that optimal carbon tax rates are 'far below the marginal environmental damages–and may even be negative' (p. 994) have imposed particular restrictions on how the proceeds are distributed (that they assert are plausible but that may be less convincing to others) and have typically not noted how sensitive results are to specifications, for instance, concerning separability of utility functions and homogeneity of preferences.

production distortions, international agencies, like the IMF, have inveighed against them and trade agreements have forced many countries to reduce import tariffs. But once we recognise that production (consumption) in the rural sector cannot be taxed, the DS analysis suggests that nonetheless imports should be taxed, and Emran and Stiglitz (2005) show that to be the case.[37]

Small open developing countries, in assessing the elasticity of supply, have to be particularly sensitive to movements of factors into and out of the country. A major factor inhibiting the taxation of capital – or the imposition of progressive taxation more generally – is that such taxes may reduce the supply of capital or of skilled labour.[38]

Finally, we observe (as we have already noted) that Ramsey assumed a well-functioning competitive economy, where the only distortions arose from taxation. In all countries – and especially in developing countries – there are a multitude of other distortions. Efficiency wages in the urban sector may give rise to unemployment. In such a case, tax policy has to take into account both the limitations on the ability to tax the rural sector and the effect of any taxation on induced unemployment.[39]

12. Concluding Comments

Ramsey's paper can be viewed as the seminal paper in what has become an immense literature on optimal taxation.[40] He focused narrowly on the question of how to minimise the deadweight losses associated with commodity taxation. His contributions – both methodological and in terms of insights into the design of tax structure – have been lasting.[41]

The subsequent literature took many turns and twists. The original suggestion, that inelastically demanded commodities like bread should face a high tax rate, was regressive enough; but Ramsey-like arguments were then used against any taxation of the income from capital, or even bequests. But even as those arguments were being formulated, deeper analyses suggested just the contrary – goods demanded by the poor

---

[37] This result has given rise to some controversy over the conditions under which a VAT can suffice. See, Keen (2008) and Boadway and Sato (2009). A key issue not addressed adequately in the literature so far is the implication of alternative tax structures for risk bearing, e.g. input taxes can increase the risk burden borne by farmers.

[38] Similar results are relevant when applying optimal tax theory to a local government within a nation, within which factors of production move freely. This severely limits the ability to impose redistributive taxes at the local level (Stiglitz, 1977, 1983a, b). This analysis has an obvious and immediate application to attempts to impose redistributive taxes within the EU.

[39] See, for instance, Sah and Stiglitz (1992) and Stiglitz (2010).

[40] There are many other directions in which Ramsey's work has had great influence. Ramsey focused on a government needing to raise a fixed amount of revenue. The amount of revenue to be raised depends on the desirable level of public goods, itself a result of balancing out the marginal benefit of public goods and the marginal costs of raising the revenue – which in turn depends on the design of the tax structure. Much of the literature already referred to address these issues. In the absence of perfect redistributive taxation, the level and form of public expenditure needs to be sensitive to distributional consequences (Stiglitz, 1998). This is a sub-theme within the huge literature in cost–benefit analysis.

[41] We have not been able to address many of the interesting and difficult questions that have been addressed in the huge literature that has grown out of Ramsey's work. For instance, natural formulations of optimal tax problems give rise to non-convexities, which in turn imply that optimal taxes should be random. See, e.g. Stiglitz (1983b), Arnott and Stiglitz (1988) and Brito *et al.* (1995). Nor we have considered the costs of administering the tax system and compliance (Slemrod, 1990).

should be taxed at lower rates, and interest income should be taxed. Though the path-breaking work of DM showed that Ramsey's results could be generalised in some important ways, other work showed that the domain of applicability of Ramsey's original insights may be more limited: changes in assumptions about the set of feasible taxes (not allowing certain taxes, or allowing a progressive income tax or non-linear commodity taxes) and, in particular, about the taxation of pure rents, incorporating more explicitly distributional considerations, and/or recognising the important ways in which our economy differs from the competitive model underlying Ramsey's analysis all change the optimal structure of commodity taxation in important ways.

Ramsey ended his 1927 essay by modestly observing that 'the more complicated results' he had obtained might 'well be valid under still wider conditions. But these are, in the general case, too complicated to be worth setting down in the absence of practical data to compare with them' (p. 60). We have noted that indeed many of the more complicated results described here require estimates of parameters about which there is little information. In the end, perhaps the most important contribution of the line of work growing out of Ramsey's seminal paper is to reinforce our basic intuitions about what kinds of tax structures make sense, remembering that if it were not for our concern about distribution – if we cared only about minimising the deadweight loss of our tax system- we would have imposed lump sum taxes: tax structures that make use of observable information, that these extensions of Ramsey's analyses support had we had the information that we would have liked to have had, taking note of the distortionary consequences of imposing taxes on these 'observable but alterable' variables. That is, to put it roughly: progressive income taxes, complemented by indirect taxation, bequest taxes and capital taxes that enhance the progressivity that can be achieved by the tax system while limiting the level of distortion (dead weight loss.)

Columbia University

References

Aiyagari, S.R. (1995). 'Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting', *Journal of Political Economy*, vol. 103(6), pp. 1158–75.
Arnott, R., Greenwald, B. and Stiglitz, J.E. (1994). 'Information and economic efficiency', *Information Economics and Policy*, vol. 6(1), pp. 77–88.
Arnott, R. and Stiglitz, J.E. (1979). 'Aggregate land rents, expenditure on public goods and optimal city size', *Quarterly Journal of Economics*, vol. 93(4), pp. 471–500.
Arnott, R. and Stiglitz, J.E. (1986). 'Moral hazard and optimal commodity taxation', *Journal of Public Economics*, vol. 29(1), pp. 1–24.
Arnott, R. and Stiglitz, J.E. (1988). 'Randomization with asymmetric information', *Rand Journal of Economics*, vol. 19(3), pp. 344–62.
Atkeson, A., Chari, V.V. and Kehoe, P.J. (1999). 'Taxing capital income: a bad idea', *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 23(3), pp. 3–17.
Atkinson, A.B. and Stiglitz, J.E. (1972). 'The structure of indirect taxation and economic efficiency', *Journal of Public Economics*, vol. 1(1), pp. 97–119.
Atkinson, A.B. and Stiglitz, J.E. (1976). 'The design of tax structure: direct versus indirect taxation', *Journal of Public Economics*, vol. 6(1), pp. 55–75.
Atkinson, A.B. and Stiglitz, J.E. (1980). *Lectures on Public Economics*, New York: McGraw-Hill Book Company.
Baumol, W.J. and Bradford, D.F. (1970). 'Optimal departures from marginal cost pricing', *American Economic Review*, vol. 60(3), pp. 265–83.

Baumol, W.J., Panzar, J.C. and Willig, R.D. (1982). *Contestable Markets and the Theory of Industry Structure*, San Diego, CA: Harcourt Brace Jovanovich.

Bevan, D. and Stiglitz, J.E. (1979). 'Intergenerational transfers and inequality', *The Greek Economic Review*, vol. 1(1), pp. 8–26.

Boadway, R. and Sato, M. (2009). 'Optimal tax design and enforcement with an informal sector', *American Economic Journal: Economic Policy*, vol. 1(1), pp. 1–27.

Boiteux, M. (1956). 'Sur la gestion des monopoles publics astreints a l'equilibre budgetaire', *Econometrica*, vol. 24(1), pp. 22–40. ('On the management of public monopolies subject to budgetary constraints', *Journal of Economic Theory*, vol. 3(3), pp. 219–40).

Bovenberg, A.L. and Goulder, L. (1996). 'Optimal environmental taxation in the presence of other taxes: general-equilibrium analyses', *American Economic Review*, vol. 86(4), pp. 985–1000.

Brito, D.L., Hamilton, J.H., Slutsky, S.M. and Stiglitz, J.E. (1990). 'Pareto efficient tax structures', *Oxford Economic Papers*, vol. 42(1), pp. 61–77.

Brito, D.L., Hamilton, J.H., Slutsky, S.M. and Stiglitz, J.E. (1995). 'Randomization in optimal income tax schedules', *Journal of Public Economics*, vol. 56(2), pp. 189–223.

Brock, W.A. (1983). 'Contestable markets and the theory of industry structure: a review article', *Journal of Political Economy*, vol. 91(6), pp. 1055–66.

Chamley, C. (1986). 'Optimal taxation of capital income in general equilibrium with infinite lives', *Econometrica*, vol. 54(3), pp. 607–22.

Chamley, C. (2001). 'Capital income taxation, wealth distribution and borrowing constraints', *Journal of Public Economics*, vol. 79(1), pp. 55–69.

Corlett, W.J. and Hague, D.C. (1953). 'Complementarity and the excess burden of taxation', *Review of Economic Studies*, vol. 21(1), pp. 21–30.

Cremer, H., Gahvari, F. and Ladoux, N. (1998). 'Externalities and optimal taxation', *Journal of Public Economics*, vol. 70(3), pp. 343–64.

Cremer, H. and Pestieau, P. (2006). 'Wealth transfer taxation: a survey of the theoretical literature', in (S.C. Kolm and J. Mercier Ythier, eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2(2), Chapter 16, pp. 1107–34, Amsterdam: North Holland Elsevier.

Cremer, H., Pestieau, P. and Rochet, J.-C. (2003). 'Capital income taxation when inherited wealth is not observable', *Journal of Public Economics*, vol. 87(11), pp. 2475–90.

Dasgupta, P. and Stiglitz, J.E. (1971). 'Differential taxation, public goods, and economic efficiency', *Review of Economic Studies*, vol. 38(2), pp. 151–74.

De Bonis, V. and Spataro, L. (2005). 'Taxing capital income as Pigouvian correction: the role of discounting the future', *Macroeconomic Dynamics*, vol. 9(4), pp. 469–77.

Diamond, P. (1975). 'A many-person Ramsey tax rule', *Journal of Public Economics*, vol. 4(4), pp. 335–42.

Diamond, P. (1978). 'Tax incidence in a two good model', *Journal of Public Economics*, vol. 9(3), pp. 283–99.

Diamond, P. and Mirrlees, J.A. (1971). 'Optimal taxation and public production', *American Economic Review*, vol. 61(1), pp. 8–27 and 261–78.

Emran, S. and Stiglitz, J.E. (2005). 'On selective indirect tax reform in developing countries', *Journal of Public Economics*, vol. 89(4), pp. 599–623.

Farhi, E. and Werning, I. (2010). 'Progressive estate taxation', *Quarterly Journal of Economics*, vol. 125(2), pp. 635–73.

Farrell, J. (1986). 'How effective is potential competition?', *Economics Letters*, vol. 20(1), pp. 67–70.

Greenwald, B. and Stiglitz, J.E. (1986). 'Externalities in economies with imperfect information and incomplete markets', *Quarterly Journal of Economics*, vol. 101(2), pp. 229–64.

Henry, G. (1886). *Progress and Poverty*, pp. 283–4. London: Kegan Paul, Tench & Co.

Harberger, A.C. (1962). 'The incidence of the corporate income tax', *Journal of Political Economy*, vol. 70(3), pp. 215–40.

Judd, K. (1985). 'Redistributive taxation in a simple perfect foresight model', *Journal of Public Economics*, vol. 28(1), pp. 59–83.

Judd, K. (1999). 'Optimal taxation and spending in general competitive growth models', *Journal of Public Economics*, vol. 71(1), pp. 1–26.

Kaplow, L. (2006). 'On the undesirability of commodity taxation even when income taxation is not optimal', *Journal of Public Economics*, vol. 90(6/7), pp. 1235–50.

Kaplow, L. (2012). 'Optimal control of externalities in the presence of income taxation', *International Economic Review*, vol. 52, pp. 487–509.

Keen, M. (2008). 'VAT, tariffs, and withholding: border taxes and informality in developing countries', *Journal of Public Economics*, vol. 92(10), pp. 1892–906.

Konishi, H. (1995). 'A Pareto-improving commodity tax reform under a smooth nonlinear income tax', *Journal of Public Economics*, vol. 56(3), pp. 413–46.

Lipsey, R.G. and Lancaster, K. (1956). 'The general theory of second best', *Review of Economic Studies*, vol. 24 (1), pp. 11–32.

Lucas, R. (1990). 'Supply-side economics: an analytical review', *Oxford Economic Papers*, New Series, vol. 42(2), pp. 293–316.

Mankiw, N.G., Weinzierl, M. and Yagan, D. (2009). 'Optimal taxation in theory and practice', Working Paper No. 15071, NBER.

Mayeres, I. and Proost, S. (1997). 'Optimal tax and public investment rules for congestion types of externalities', *Scandinavian Journal of Economics*, vol. 99(2), pp. 261–79.

Meade, J.E. (1955). *The Theory of International Economic Policy. Volume II: Trade and Welfare*, Oxford: Oxford University Press.

Mieszkowski, P. (1969). 'Tax incidence theory', *Journal of Economic Literature*, vol. 7(4), pp. 1103–24.

Mirrlees, J.A. (1971). 'An exploration in the theory of optimal income taxation', *Review of Economic Studies*, vol. 38(2), pp. 175–208.

Mirrlees, J.A. (1972). 'On producer taxation', *Review of Economic Studies*, vol. 39(1), pp. 105–11.

Mirrlees, J.A. (1975). 'Optimal commodity taxation in a two-class economy', *Journal of Public Economics*, vol. 4(1), pp. 27–33.

Mirrlees, J.A. (1976). 'Optimal tax theory: a synthesis', *Journal of Public Economics*, vol. 6(4), pp. 327–58.

Naito, H. (1999). 'Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency', *Journal of Public Economics*, vol. 71(2), pp. 165–88.

Nordhaus, W.D. (1993). 'Optimal greenhouse gas reductions and tax policy in the "DICE" model', *American Economic Review*, vol. 83(2), pp. 313–7.

Parry, I. (1994). 'Pollution taxes and revenue recycling', *Journal of Environmental Economics and Management*, vol. 29(3), pp. S64–77.

Pigou, A.C. (1928). *A Study in Public Finance*, London: Macmillan and Co.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Cambridge, MA: Belknap (Harvard University) Press.

Piketty, T. and Saez, E. (2012). 'A theory of optimal capital taxation', Working Paper No. 17989, NBER.

Piketty, T., Saez, E. and Stantcheva, S. (2011). 'Taxing the 1%: why the top tax rate could be over 80%', *Vox*, December 8.

Ramsey, F.P. (1926). 'Truth and probability', Published 1931 in (R.B. Braithwaite, ed.), *Foundations of Mathematics and other Logical Essays*, Ch. VII, pp. 156–98, London: Kegan, Paul, Trench, Trubner & Co. Ltd; New York: Harcourt, Brace and Company.

Ramsey, F.P. (1927). 'A contribution to the theory of Taxation', ECONOMIC JOURNAL, vol. 37(145), pp. 47–61.

Ramsey, F.P. (1928). 'A mathematical theory of saving', ECONOMIC JOURNAL, vol. 38(152), pp. 543–59.

Saez, E. (2002). 'The desirability of commodity taxation under nonlinear income taxation and heterogeneous tastes', *Journal of Public Economics*, vol. 83(2), pp. 217–30.

Sah, R. and Stiglitz, J.E. (1992). *Peasants versus City-dwellers: Taxation and the Burden of Economic Development*. Oxford: Oxford University Press.

Salanié, B. (2003). *The Economics of Taxation*, Cambridge, MA.: MIT Press.

Sandmo, A. (1975). 'Optimal tazation in the presence of externalities', *Swedish Journal of Economics*, vol. 77(1), pp. 86–98.

Sandmo, A. (1976). 'Direct vs. indirect Pigovian taxation', *European Economic Review*, vol. 7(4), pp. 337–49.

Sappington, D. and Stiglitz, J.E. (1987*a*). 'Information and regulation', in (E. Bailey, ed.), *Public Regulation*, pp. 3–43, London: MIT Press.

Sappington, D. and Stiglitz, J.E. (1987*b*). 'Privatization, information and incentives', *Journal of Policy Analysis and Management*, vol. 6(4), pp. 567–82.

Slemrod, J. (1990). 'Optimal taxation and optimal tax systems', *Journal of Economic Perspectives*, vol. 4(1), pp. 157–78.

Stiglitz, J.E. (1969). 'The effects of income, wealth and capital gains taxation on risk-taking', *Quarterly Journal of Economics*, vol. 83(2), pp. 263–83.

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, New York: Praeger.

Stiglitz, J.E. (1976). 'Notes on estate taxes, redistribution, and the concept of balanced growth path incidence', *Journal of Political Economy*, vol. 86(2), pp. 137–50.

Stiglitz, J.E. (1977). 'Theory of local public goods', in (M.S. Feldstein and R.P. Inman, eds.), *The Economics of Public Service*, pp. 274–333, London: MacMillan Press.

Stiglitz, J.E. (1978). 'Equality, taxation and inheritance', in (W. Krelle and A.F. Shorrocks, eds.), *Personal Income Distribution*, pp. 271–303, Amsterdam: North Holland.

Stiglitz, J.E. (1982*a*). 'Self-selection and Pareto efficient taxation', *Journal of Public Economics*, vol. 17(2), pp. 213–40.

Stiglitz, J.E. (1982*b*). 'The rate of discount for cost-benefit analysis and the theory of the second best', in (R. Lind, ed.), *Discounting for Time and Risk in Energy Policy*, pp. 151–204, Washington, DC: Resources for the Future.

Stiglitz, J.E. (1983*a*). 'Public goods in open economies with heterogeneous individuals', in (J.F. Thisse and H.G. Zoller, eds.), *Locational Analysis of Public Facilities*, pp. 55–78, Amsterdam: North-Holland.

Stiglitz, J.E. (1983*b*). 'The theory of local public goods twenty-five years after Tiebout: a perspective', in (G.R. Zodrow, ed.), *Local Provision of Public Services: The Tiebout Model After Twenty-Five Years*, pp. 17–53, New York: Academic Press.

Stiglitz, J.E. (1985). 'Inequality and capital taxation', IMSSS Technical Report 457, Stanford University.

Stiglitz, J.E. (1987). 'Pareto efficient and optimal taxation and the new new welfare economics', in (A. Auerbach and M. Feldstein, eds.), *Handbook of Public Economics, 2*, pp. 991–1042, Amsterdam: North Holland.

Stiglitz, J.E. (1988). 'Technological change, sunk costs, and competition', in (M.N. Baily and C. Winston, eds.), *Special Issue on Macroeconomics*, pp. 883–947, Washington, DC: Brookings Institution.

Stiglitz, J.E. (1998). 'Pareto efficient taxation and expenditure policies, with applications to the taxation of capital, public investment, and externalities', presented at conference in honor of Agnar Sandmo, Bergen, January. (Revision and extension of Stiglitz (1985)).

Stiglitz, J.E. (2009). 'Simple formulae for optimal income taxation and the measurement of inequality', in (K. Basu and R. Kanbur, eds.), *Arguments for a Better World: Essays in Honor of Amartya Sen, Volume I, Ethics, Welfare, and Measurement*, pp. 535–66, Oxford: Oxford University Press.

Stiglitz, J.E. (2010). 'Development-oriented tax policy', in (R.H. Gordon, ed.), *Taxation in Developing Countries*, pp. 11–36, New York: Columbia University Press.

Stiglitz, J.E. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future*. New York: W.W. Norton.

Stiglitz, J.E. (2013). 'Sharing the burden of saving the planet: global social justice for sustainable development', in (M. Kaldor and J.E. Stiglitz, eds.), *The Quest for Security: Protection without Protectionism and the Challenge of Global Governance*, pp. 161–90, New York: Columbia Unversity Press.

Stiglitz, J.E. (2014). 'Learning, growth, and development: a lecture in honor of Sir Partha Dasgupta', in (S. Barret, K-G. Maler and E.S. Maskin, eds.), *Environment & Development Economics: Essays in Honor of Sir Partha Dasgupta*, pp. 31–94, Oxford: Oxford University Press.

Vickrey, W. (1977). 'The city as a firm', in (M.S. Feldstein and R.P. Inman, eds.), *The Economics of Public Services*, pp. 334–43, London: Macmillan.

Vickrey, W. (1999). 'Henry George, economies of scale, and land taxation', in (K.C. Wenzer, ed.), *Land Value Taxation*, pp. 24–37, New York: M. E. Sharpe.

Appendix A. Ramsey, F.P. (1927). 'A contribution to the theory of taxation', ECONOMIC JOURNAL, vol. 37(145), pp. 47–61.

A CONTRIBUTION TO THE THEORY OF TAXATION

THE problem I propose to tackle is this : a given revenue is to be raised by proportionate taxes on some or all uses of income, the taxes on different uses being possibly at different rates; how should these rates be adjusted in order that the decrement of utility may be a minimum? I propose to neglect altogether questions of distribution and considerations arising from the differences in the marginal utility of money to different people; and I shall deal only with a purely competitive system with no foreign trade. Further I shall suppose that, in Professor Pigou's terminology, private and social net products are always equal or have been made so by State interference not included in the taxation we are considering. I thus exclude the case discussed in Marshall's *Principles* in which a bounty on increasing-return commodities is advisable. Nevertheless we shall find that the obvious solution that there should be no differentiation is entirely erroneous.

The effect of taxation is to transfer income in the first place from individuals to the State and then, in part, back again to rentiers and pensioners. These transfers will slightly alter the demand schedules in a way depending on the incidence of the taxes and the manner of their expenditure. I neglect these alterations;[1] and I also suppose that " a given revenue " means a given money revenue, "money" being so adjusted that its marginal utility is constant.

This problem was suggested to me by Professor Pigou, to whom I am also indebted for help and encouragement in its solution.

In the first part I deal with the perfectly general utility function and establish a result which is valid for a sufficiently small revenue, and takes a peculiarly simple form if we can treat the revenue as an infinitesimal. I prove, in fact, that in raising an infinitesimal revenue by proportionate taxes on given commodities the taxes should be such as to diminish in the same proportion the production of each commodity taxed.

In the second part I assume that the utility function is quadratic, which means roughly that the supply and demand

---

[1] The outline of a more general treatment is given in the Appendix.

curves are straight lines, but does not exclude the most general possibilities of joint supply and joint demand. With this assumption we can show that the rule given above for an infinitesimal revenue is valid for any revenue which can be raised at all.

In the third part I give certain important special cases of these general theorems; and in part four indicate certain practical applications.

Part I

(1) I suppose there to be altogether n commodities on which incomes are spent and denote the quantities of them which are produced in a unit of time by $x_1, x_2 \ldots x_n$. Some of these commodities may be identical, save for the place or manner of their production or consumption; *e.g.*, we can regard sugar used in tea as a different commodity from sugar used in coffee, and corn grown in Norfolk as different from that grown in Suffolk. In order to avoid double reckoning we suppose that these commodities are all either consumed or saved; *e.g.*, we include household coal, but not industrial coal except in so far as an increase in the stock of industrial coal is a form of saving, so that this rate of *increase* can form one of our quantities x. The quantities $x_1, x_2 \ldots$ can be measured in any convenient different units.

(2) We denote by $u = F(x_1 \ldots x_n)$ the *net* utility of producing and consuming (or saving) these quantities of commodities. This is usually regarded as the difference of two functions, one of which represents the utility of consuming, the other the disutility of producing. But so to regard it is to make an unnecessary assumption of independence between consumption and production; to assume, for instance, that the utility of a hot bath is the same whether one does or does not work in a coal mine. This assumption we do not require to make.

(3) If there is no taxation stable equilibrium will occur for values of the x's which make u a maximum. Let us call these values $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_n$ or collectively the point P. Then at P we have

$$\frac{\partial u}{\partial x_r} = 0 \qquad\qquad r = 1, \ldots n.$$

$d^2u = \Sigma\Sigma \dfrac{\partial^2 u}{\partial x_r \partial x_s} dx_r dx_s$ is a negative definite form.

Suppose now taxes are levied on the different commodities

at the rates λ_1, λ_2 . . . λ_n per unit in money whose marginal utility is unity. Then the new equilibrium is determined by

$$\frac{\partial u}{\partial x_r} = \lambda_r{}^1 \qquad r = 1, \ldots n \quad . \quad . \quad . \quad . \quad (1)$$

In virtue of these equations we can regard the λ's as functions of the x's, which vanish at P, and satisfy identically

$$\frac{\partial \lambda_r}{\partial x_s} = \frac{\partial \lambda_s}{\partial x_r} \cdot \left(= \frac{\partial^2 u}{\partial x_r \partial x_s} \right) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Also the revenue $R = \Sigma \lambda_r x_r$.

We shall always suppose R to be positive, but there is no *a priori* reason why some of the λ's should not be negative; they will then, of course, represent bounties.

(4) Our first problem is this : given R, how should the λ's be chosen in order that the values of the x's given by equations (1) shall make u a maximum.

I.e., u is to be a maximum subject to $\Sigma \lambda_r x_r = R \left(\text{where } \lambda_r \text{ is } \frac{\partial u}{\partial x_r} \right)$.

We must have

$$0 = du = \Sigma \lambda_r dx_r \text{ for any values of } dx_r$$

subject to

$$0 = dR = \Sigma \lambda_r dx_r + \underset{r \; s}{\Sigma \Sigma} x_s \frac{\partial \lambda_s}{\partial x_r} dx_r,$$

and so we have

$$\frac{\lambda_1}{\underset{s}{\Sigma} x_s \frac{\partial \lambda_s}{\partial x_1}} = \frac{\lambda_2}{\underset{s}{\Sigma} x_s \frac{\partial \lambda_s}{\partial x_2}} = \qquad = \qquad = \frac{\lambda_n}{\underset{s}{\Sigma} x_s \frac{\partial \lambda_s}{\partial x_n}} \quad . \quad . \quad (3)$$

$$= \frac{R}{\Sigma \Sigma \frac{\partial \lambda_s}{\partial x} x_r x_s} = - \theta \text{ (say)}.$$

(5) These equations determine values of the x's which are critical for u, and it remains to discuss the possibility of a plurality of solutions and to determine conditions under which they give a true maximum. We shall show that if R is small enough they will have a unique solution $x_1, x_2 \ldots x_n$, which tends to $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_n$ as $R \to 0$, and that this solution will make u a true maximum.

---

[1] *E.g.*, if $u = u_1 - u_2$ (consumers' utility − producers' disutility)

$\frac{\partial u}{\partial x_r} = \frac{\partial u_1}{\partial x_r} - \frac{\partial u_2}{\partial x_r}$ = demand price of rth commodity − supply price = tax.

No. 145.—VOL. XXXVII.   E

For, since $d^2u = \Sigma\Sigma \dfrac{\partial \lambda_r}{\partial x_s} dx_r dx_s$ is negative definite at P,

$(-)^n \dfrac{\partial(\lambda_1, \lambda_2 \ldots \lambda_n)}{\partial(x_1, x_2 \ldots x_n)}$ is positive at, and therefore near, P. Hence we can express the x's as functions of the λ's. The equations (3) then become

$$\lambda_r = R\psi_r(\lambda_1, \ldots \lambda_n) \qquad r = 1, 2, \ldots n.$$

For the denominator $\Sigma\Sigma \dfrac{\partial \lambda_s}{\partial x_r} x_r x_s$ is a negative definite form with d^2u and so cannot vanish near P (and therefore also $\theta > 0$). The Jacobian of these last equations with regard to the λ's will tend to 1 as R tends to 0, and they will therefore have a unique solution $\lambda_1, \ldots \lambda_n$ which tends to 0, 0 $\ldots$ 0 as R tends to 0. Hence the equations (3) have a unique solution tending to P as $R \to 0$.

We have now to consider the conditions for a maximum which are obtained most simply by Lagrange's multipliers.

If we consider $u + KR$

we should have $\qquad \dfrac{\partial u}{\partial x_r} + K\dfrac{\partial R}{\partial x_r} = 0$

or $\quad 1 + K - \dfrac{K}{\theta} = 0 \quad$ if θ has the meaning it has in equations (3).

or $\qquad K = \dfrac{+\theta}{1-\theta}.$

Then $\qquad d^2u = d^2\Big(u + \dfrac{\theta}{1-\theta}R\Big)$

$$= d^2u + \dfrac{\theta}{1-\theta}d^2R$$

(calculated as if the variables x were independent [1]), and in a sufficiently small neighbourhood of P we shall have $\theta <$ any assigned positive constant and so $d^2u + \dfrac{\theta}{1-\theta}d^2R$ negative definite with d^2u. This establishes the desired result.[2]

(6) Suppose now R and the λ's can be regarded as infinitesimals; then putting

$$\lambda_r = \Sigma_s \dfrac{\partial \lambda_r}{\partial x_s} dx_s$$

equations (3) give us, using (2),

[1] See, e.g., de la Vallée Poussin, *Cours d'Analyse*, 4th ed., t. 1, p. 149.

[2] Clearly also we shall get a maximum at any point for which d^2R is negative and $\theta < 1$; i.e., if d^2R is everywhere negative (3) will give a maximum for all values of θ up to $\theta = 1$, which gives a maximum of R. This covers the case treated in Part II and so also any case approximating to that.

$$\frac{\underset{s}{\Sigma}\frac{\partial \lambda_r}{\partial x_s}dx_s}{\underset{s}{\Sigma}\frac{\partial \lambda_r}{\partial x_s}x_s} = \qquad = \qquad = -\theta = \frac{R}{\Sigma\Sigma x_r x_s \frac{\partial \lambda_r}{\partial x_s}} < 0,$$

and their solution is evidently given by

$$\frac{dx_1}{x_1} = \frac{dx_2}{x_2} = \qquad = \frac{dx_n}{x_n} = -\theta < 0 \quad . \quad . \quad . \quad (4)$$

i.e., the production of each commodity should be diminished in the same proportion.

(7) It is interesting to extend these results to the case of a given revenue to be raised by taxing certain commodities only. If the utility were the sum of two functions, one of the taxed and the other of the untaxed commodities, it is obvious that our conclusions would be the same as before. But in the general case the question is by no means so simple.

Let us denote the quantities of the commodities to be taxed by $x_1 \ldots x_n$, and those not to be taxed by $y_1 \ldots y_n$.

If $\quad \lambda_r = \dfrac{\partial u}{\partial x_r} \quad$ then λ_r is the tax per unit on x_r,

and if $\quad \mu_r = \dfrac{\partial u}{\partial y_r}, \ \mu_r = 0 \quad$ (λ's and μ's functions of x's and y's),

also as before

$$\frac{\partial \lambda_r}{\partial x_s} = \frac{\partial \lambda_s}{\partial x_r}, \ \frac{\partial \mu_r}{\partial y_s} = \frac{\partial \mu_s}{\partial y_r}, \text{ and } \frac{\partial \lambda_r}{\partial y_s} = \frac{\partial \mu_s}{\partial x_r} \quad . \quad . \quad . \quad (5)$$

and we have to maximise u subject to

$$\underset{r=1}{\Sigma^n}\lambda_r x_r = R, \ \mu_t = 0, \ t = 1, \ldots m.$$

We have

$$0 = du = \underset{r}{\Sigma}\lambda_r dx_r$$

$$0 = dR = \underset{r}{\Sigma}\lambda_r dx_r + \underset{s}{\Sigma}\underset{r}{\Sigma}x_s \frac{\partial \lambda_s}{\partial x_r}dx_r + \underset{s}{\Sigma}\underset{t}{\Sigma}x_s \frac{\partial \lambda_s}{\partial y_t}dy_t$$

$$0 = d\mu_t = \underset{r}{\Sigma}\frac{\partial \mu_t}{\partial x_r}dx_r + \underset{u}{\Sigma}\frac{\partial \mu_t}{\partial y_u}dy_u, \quad t = 1, \ldots m.$$

Solving these last equations ($d\mu_t = 0$) for the dy's we obtain

$$dy_t = \underset{r}{\Sigma}\chi_{tr}dx_r \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $\quad \dfrac{\partial \mu_t}{\partial x_r} + \underset{u=1}{\Sigma^m}\dfrac{\partial \mu_t}{\partial y_u}\chi_{ur} = 0 \ \begin{Bmatrix} r = 1, \ldots n \\ t = 1, \ldots m \end{Bmatrix} . \quad . \quad . \quad (7)$

(The possibility of solution is guaranteed by the discriminants of d^2u not vanishing.)

Whence $\quad 0 = dR = \underset{r}{\Sigma}dx_r\Big(\lambda_r + \underset{s}{\Sigma}x_s\frac{\partial \lambda_s}{\partial x_r} + \underset{s}{\Sigma}\underset{t}{\Sigma}x_s\frac{\partial \lambda_s}{\partial y_t}\chi_{tr}\Big).$

E 2

∴ instead of equations (3) we have

$$\frac{\lambda_r}{\sum\limits_{s=1}^{n} x_s \left(\frac{\partial \lambda_s}{\partial x_r} + \sum\limits_{t=1}^{m} \chi_{tr} \frac{\partial \lambda_s}{\partial y_t}\right)} = \qquad = \qquad . \quad . \quad (3')$$

It can be shown that these give a maximum of u with the same sort of limitations as equations (3) do.

(8) And if the λ's are infinitesimal

$$\lambda_r = \sum_s \frac{\partial \lambda_r}{\partial x_s} dx_s + \sum_t \frac{\partial \lambda_r}{\partial y_t} dy_t$$

$$= \sum_{s=1}^{n} dx_s \left(\frac{\partial \lambda_s}{\partial x_r} + \sum_{t=1}^{m} \chi_{ts} \frac{\partial \mu_t}{\partial x_r}\right) \quad \text{by (5), (6).}$$

But $\sum\limits_t \chi_{ts} \dfrac{\partial \mu_t}{\partial x_r} = - \sum\limits_t \sum\limits_u \dfrac{\partial \mu_t}{\partial y_u} \chi_{ts} \chi_{ur}$ by (7)

$$= \sum_t \chi_{tr} \frac{\partial \mu_t}{\partial x_s} \left(\text{by symmetry since } \frac{\partial \mu_t}{\partial y_u} = \frac{\partial \mu_u}{\partial y_t}\right).$$

So $\lambda_r = \sum\limits_s dx_s \left(\dfrac{\partial \lambda_s}{\partial x_r} + \sum\limits_t \chi_{tr} \dfrac{\partial \lambda_s}{\partial y_t}\right)$, since $\dfrac{\partial \lambda_s}{\partial y_t} = \dfrac{\partial \mu_t}{\partial x_s}$,

and so equations (3′) are satisfied by

$$\frac{dx_1}{x_1} = \qquad = \frac{dx_n}{x_n},$$

i.e., as before the taxes should be such as to reduce in the same proportion the production of each *taxed* commodity.

(9) Further than this it is difficult to go without making some new assumption. The assumption I propose is perhaps unnecessarily restrictive, but it still allows scope for all possible first-order relations between commodities in respect of joint supply or joint demand, and it has the great merit of rendering the problem completely soluble.

I shall assume that the utility is a non-homogeneous quadratic function of the x's, or that the λ's are linear. This assumption simplifies the problem in precisely the same way as we have previously simplified it by supposing the taxes to be infinitesimal. We shall, however, make this new assumption the occasion for exhibiting a method of interpreting our formulae geometrically in a manner which makes their meaning and mutual relations considerably clearer.

It is not, of course, necessary, nor would it be sensible to suppose the utility function quadratic for all values of the variables; we need only suppose it so for a certain range of values round the point P, such that there is no question of imposing taxes large enough to move the production point (values of the

x's) outside this range. If we were concerned with independent commodities, this assumption would mean that the taxes were small enough for us to treat the supply and demand curves as straight lines.

PART II

(10) Let $u = \text{constant} + \Sigma a_r x_r + \Sigma\Sigma\beta_{rs}x_r x_s$, $(\beta_{rs} = \beta_{sr})$, and let us regard the x's as rectangular Cartesian co-ordinates of points in n-dimensional space.

The point $P(\bar{x}_1, \ldots \bar{x}_n)$ is given by $\dfrac{\partial u}{\partial x_r} = 0$,

and at that point

$d^2u \doteqdot 2\Sigma\Sigma\beta_{rs}dx_r dx_s$ is a negative definite form.

$\therefore \Sigma\Sigma\beta_{rs}x_r x_s$ is a negative definite form,

and the loci $u = \text{constant}$ are hyper-ellipsoids with the point P for centre.

Since
$$\lambda_r = \frac{\partial u}{\partial x_r} = a_r + 2\Sigma_s\beta_{rs}x_s \quad \ldots \quad \ldots \quad (8)$$

$$R = \Sigma\lambda_r x_r = \Sigma a_r x_r + 2\Sigma\Sigma\beta_{rs}x_r x_s \quad \ldots \quad (9)$$

and the loci $R = \text{constant}$ are hyper-ellipsoids with the point Q, whose co-ordinates are $\frac{1}{2}\bar{x}_1, \frac{1}{2}\bar{x}_2 \ldots \frac{1}{2}\bar{x}_n$, for centre.

(The equations for Q are those for P with their first degree terms doubled and their constant terms unaltered.)

Moreover, the hyper-ellipsoids $u = \text{constant}$, $R = \text{constant}$ are all similar and similarly situated. The figure shows these relations for the case of two commodities only.
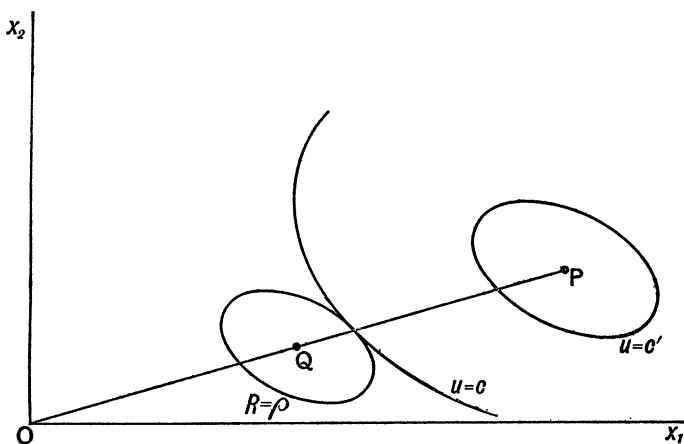
(11) If we are to raise a revenue ρ we must depress production to some point on the hyper-ellipsoid $R = \rho$.[1]

[1] We can depress production to any point we please because the connection between the x's and λ's is one-one.

To do this so as to make u a maximum we must choose a point on this hyper-ellipsoid at which it touches an ellipsoid of the family $u = $ constant. There will be two such points which will lie on the line PQ : one between Q and P making u a maximum, the other between O and Q making u a minimum. For the point of contact of two similar and similarly situated hyper-ellipsoids must lie on the line joining their centres. Since the maximum of u is given by a point on OP we have as before that

The taxes should be such as to diminish the production of all commodities in the same proportion.

And this result is now valid not merely for an infinitesimal revenue but for any revenue which it is possible to raise at all.

The maximum revenue will be obtained by diminishing the production of each commodity to one-half of its previous amount, *i.e.*, to the point Q.

(12) If in accordance with this rule we impose taxes reducing production from $\bar{x}_1,\ \bar{x}_2 \ldots \bar{x}_n$ to $(1 - k)\bar{x}_1,\ (1 - k)\bar{x}_2 \ldots (1 - k)\bar{x}_n$.

We get from (8) $\lambda_r = a_r + 2(1 - k)\underset{s}{\Sigma}\beta_{rs}\bar{x}_r$,

but at P $\qquad\qquad \lambda_r = 0$, so that $0 = a_r + 2\underset{r}{\Sigma}\beta_{rs}\bar{x}_r$;

therefore $\qquad\qquad \lambda_r = ka_r$ . . . . (10)

i.e., the taxes should be in the fixed proportions $\lambda_1 : \lambda_2 :$ $: \lambda_n : : a_1 : a_2 \qquad : a_n$ independent of the revenue to be raised.

Also $R = \Sigma\lambda_r x_r = k(1 - k)\Sigma a_r\bar{x}_r$,
$\qquad = 4k(1 - k) \times$ the maximum revenue (got by putting $k = \frac{1}{2}$).

(13) Since k is positive it follows from (10) that the sign of λ_r is the same as that of a_r, and unless the a_r are all positive some of the λ_r will be negative, and the most expedient way of raising a revenue will be by placing bounties on some commodities and taxes on others.

The sort of case in which this might occur is that of sugar and particularly sour fruits, *e.g.* damsons. A tax on sugar might reduce the consumption of damsons more than in proportion to the reduction in the total consumption of sugar and so require to be offset by a bounty on damsons.

(14) We can now consider the more general problem : a given revenue is to be raised by means of fixed taxes $\mu_1 \ldots \mu_m$ on m commodities and by taxes to be chosen at discretion on the remainder. How should they be chosen in order that utility may be a maximum?

We have $\lambda_1 = \mu_1, \ldots \lambda_n = \mu_m$, m hyperplanes $(n - 1$ folds) whose intersection is a plane $n - m$ fold which we will call S. S will cut the hyper-ellipsoids $u = $ constant, $R = $ constant in hyper-ellipsoids which are similar and similarly situated and whose centres are the points P', and Q' in which S is met by the m-folds through P and Q conjugate to S in $u = c$ or $R = c$. As before the required maximum is given by the point of contact of two of these hyper-ellipsoids in S, which must lie upon the line $P'Q'$.

Now the hyperplane $\lambda_1 = \mu_1$ or $\dfrac{\partial u}{\partial x_1} = \mu_1$ is conjugate in $u = c$ to the diameter

$$x_2 = \bar{x}_2, \; x_3 = \bar{x}_3, \; \ldots \; x_n = \bar{x}_n.$$

Hence S is conjugate to the m-fold

$$x_{m+1} = \bar{x}_{m+}, \; \ldots, \; x_n = \bar{x}_n,$$

and the co-ordinates of P' satisfy these equations, since they lie on this m-fold.

Similarly the co-ordinates of Q' satisfy

$$x_{m+1} = \tfrac{1}{2}\bar{x}_{m+1}, \; \ldots \; x_n = \tfrac{1}{2}\bar{x}_n.$$

And so the desired production point lying on the line $P'Q'$ satisfies

$$\frac{x_{m+1}}{\bar{x}_{m+1}} = \frac{x_{m+2}}{\bar{x}_{m+2}} = \qquad = \frac{x_n}{\bar{x}_n},$$

i.e., the whole system of taxes must be such as to reduce in the same proportion the production of the commodities taxed at discretion.

PART III

(15) I propose now to explain what our results reduce to in certain special cases. First suppose that all the commodities are independent and have their own supply and demand equations, *i.e.*, we have for the rth commodity the demand price

$$p_r = \phi_r(x_r)$$

and the supply price
$$q_r = f_r(x_r).$$

$\therefore$
$$\lambda_r = p_r - q_r = \phi_r(x_r) - f_r(x_r),$$

and equations (3) become, since $\dfrac{\partial \lambda_r}{\partial x_s} = 0$, $r \neq s$,

$$\frac{\lambda_1}{x_1\{\phi_1{}'(x_1) - f_1{}'(x_1)\}} = \frac{\lambda_2}{x_2\{\phi_2{}'(x_2) - f_2{}'(x_2)\}} = \ldots = -\theta.$$

These equations we can express in terms of elasticities in the following way.

Suppose the tax *ad valorem* (reckoned on the price got by the producer) on the rth commodity is μ_r, then

$$\lambda_r = \mu_r q_r = \mu_r f_r(x_r),$$

and $\phi_r(x_r) = f_r(x_r) + \lambda_r = (1 + \mu_r)f_r(x_r).$

$\therefore$ $\theta = \dfrac{-\lambda_r}{x_r\{\phi_r{}'(x_r) - f_r{}'(x_r)\}} = \dfrac{+\mu_r}{x_r\dfrac{f_r{}'(x_r)}{f_r(x_r)} - (1 + \mu_r)x_r\dfrac{\phi_r{}'(x_r)}{\phi_r(x_r)}},$

now $x_r\dfrac{f_r{}'(x_r)}{f_r(x_r)}$ is the reciprocal of the elasticity of supply of the commodity reckoned positive for diminishing returns, and $-x_r\dfrac{\phi_r{}'(x_r)}{\phi_r(x_r)}$ is the reciprocal of the elasticity of demand, reckoned positive in the normal case.

Hence if we denote by ρ_r and ϵ_r the elasticities of demand and supply,

$$\mu_r = \theta\left(\frac{1}{\epsilon_r} + \frac{1 + \mu_r}{\rho_r}\right),$$

or $$\mu_r = \frac{\left(\dfrac{1}{\epsilon_r} + \dfrac{1}{\rho_r}\right)\theta}{1 - \dfrac{\theta}{\rho_r}} \qquad . \quad . \quad . \quad . \quad . \quad (11)$$

(valid provided the revenue is small enough, see § 5).

For infinitesimal taxes θ is infinitesimal and

$$\frac{\mu_1}{\dfrac{1}{\epsilon_1} + \dfrac{1}{\rho_1}} = \qquad = \qquad = \frac{\mu_n}{\dfrac{1}{\epsilon_n} + \dfrac{1}{\rho_n}} \quad . \quad . \quad (12)$$

i.e., the tax *ad valorem* on each commodity should be proportional to the sum of the reciprocals of its supply and demand elasticities.

(16) It is easy to see

(1) that the same rule (12) applies if the revenue is to be collected off certain commodities only, which have supply and demand schedules independent of each other and all other commodities, even when the other commodities are not independent of one another.

(2) The rule does not justify any bounties; for in stable equilibrium, although $\dfrac{1}{\epsilon_r}$ may be negative, $\dfrac{1}{\rho_r} + \dfrac{1}{\epsilon_r}$ must be positive.

(3) If any one commodity is absolutely inelastic, either for supply or for demand, the whole of the revenue should be

collected off it. This is independently obvious, for taxing such a commodity does not diminish utility at all. If there are several such commodities the whole revenue should be collected off them, it does not matter in what proportions.

(17) Let us next take the case in which all the commodities have independent demand schedules but are complete substitutes for supply; *i.e.*, with appropriate units the demand price

$$p_r = \phi_r(x_r),$$

the supply price $\qquad q_r = f(x_1 + \quad + x_n).$

Let us put $\qquad z = x_1 + \quad + x_n.$

We can imagine this case as that of a country in which all commodities are produced at constant returns by the application of one kind of labour only, the increase in the supply price arising solely from the increasing marginal disutility of labour, and the commodities satisfying independent needs. Then z will represent the amount of labour.

Equations (3) give us

$$- \theta = \frac{\lambda_r}{x_r\phi_r'(x_r) - zf'(z)}.$$

Or if μ_r represents the tax *ad valorem* and ρ_r the elasticity of demand for the rth commodity and ϵ the elasticity of supply of things in general, we get, by a similar process to that of § 15,

$$\mu_r = \frac{\left(\dfrac{1}{\rho_r} + \dfrac{1}{\epsilon}\right)\theta}{1 - \dfrac{\theta}{\rho_r}} . \quad . \quad . \quad . \quad . \quad (13)$$

If the taxes are infinitesimal we have

$$\frac{\mu_r}{\dfrac{1}{\rho_r} + \dfrac{1}{\epsilon}} = \quad = \quad = \theta . \quad . \quad . \quad . \quad (14)$$

In this case we see that if the supply of labour is fixed (absolutely inelastic, $\epsilon \to 0$) the taxes should be at the same *ad valorem* rate on all commodities.

(19) If some commodities only are to be taxed it is easier to work from the result proved in § 8 for an infinitesimal revenue, that the production of the commodities taxed should be diminished in the same ratio.

Suppose, then, $x_1, \ldots x_m$ are to be taxed, $x_{m+1} \ldots x_n$ untaxed.

Let $\qquad dx_1 = - kx_1, \ldots, dx_m = - kx_m.$

Let
$$z' = x_1 + x_2 + \ldots + x_m$$
$$z'' = x_{m+1} + \ldots + x_n.$$
$$\lambda_1 = \phi_1(x + dx_1) - zf(z + dz)$$
$$= \phi_1'(x_1)dx_1 - f'(z)dz.$$

$\therefore$
$$\mu_1 = \frac{k}{\rho_1} - \frac{dz}{z\epsilon} \ldots, \mu_m = \frac{k}{\rho_m} - \frac{dz}{\epsilon z},$$

now
$$dz = dz' + dz'' = -kz' + dz'',$$

also
$$0 = -\frac{dx_{m+1}}{\rho_{m+1}x_{m+1}} - \frac{dz}{\epsilon z}.$$

$\therefore$
$$\frac{dx_{m+1}}{\rho_{m+1}x_{m+1}} = \frac{dx_{m+2}}{\rho_{m+2}x_{m+2}} = \frac{dx_n}{\rho_n x_n} = -\frac{dz}{\epsilon z} = \frac{dz''}{\sum_{m+1}^{n}\rho_r x_r} = \frac{kz'}{\epsilon z + \sum_{m+1}^{n}\rho_r x_r}.$$

$\therefore$
$$\mu_1 = k\left(\frac{1}{\rho_1} + \frac{\sum_1^m x_r}{\epsilon\sum_1^n x_r + \sum_{m+1}^n \rho_r x_r}\right), \text{ etc.}$$

As before we see that of two commodities that should be taxed most which has the least elasticity of demand, but that if the supply of labour is absolutely inelastic all the commodities should be taxed equally.

PART IV

(20) We come now to applications of our theory; these cannot be made at all exactly without data which I, at any rate, do not possess. The simplest result is the one which we have proved in the general case for an infinitesimal revenue (§ 8); this means that it is approximately true for small revenues, and that the approximation approaches perfection as the revenue approaches zero. It is thus logically similar to the theorem that the period of oscillation of a pendulum is independent of the amplitude. We have also extended the result to any revenue which does not take the production point outside a region in which the utility may be taken to be quadratic, *i.e.*, the supply and demand schedules linear.

The sort of cases in which our theory may be useful are the following :

(21) (*a*) If a commodity is produced by several different methods or in several different places between which there is no mobility of resources, it is shown that it will be advantageous to discriminate between them and tax most the source of supply which is least elastic. For this will be necessary if we are to maintain unchanged the proportion of production between the two sources (result analogous to § 19 with supply and demand interchanged).

(b) If several commodities which are independent for demand require precisely the same resources for their production, that should be taxed most for which the elasticity of demand is least (§ 19).

(c) In taxing commodities which are rivals for demand, like wine, beer and spirits, or complementary like tea and sugar, the rule to be observed is that the taxes should be such as to leave unaltered the proportions in which they are consumed (§ 14). Whether the present taxes satisfy this criterion I do not know.

(d) In the case of the motor taxes we must separate off so much of the taxation as is offset by damage to the roads. This part should be so far as possible equal to the damage done. The remainder is a genuine tax and should be distributed according to our theory; that is to say, it should be placed partly on petrol and partly on motor-cars, so as to preserve unchanged the proportion between their consumption, and should be distributed between Fords and Morrises, so as to reduce their output in the same ratio. The present system fails in both these respects.

(22) (e) Another possible application of our theory is to the question of exempting savings from income-tax.[1] We may consider two uses of income only, saving and spending, and supposing them independent we may use the result (13) in § 17. We must suppose the taxes imposed only for a very short time [2] and that they raise no expectation of similar taxation in the future; since otherwise we require a mathematical theory considerably more difficult than anything in this paper.

On these assumptions, since the amount of saving in the very short time cannot be sufficient to alter appreciably the marginal utility of capital, the elasticity of demand for saving will be infinite, and we have

$$\mu_1 \text{ (tax on spending)} = \frac{\left(\dfrac{1}{\rho_1} + \dfrac{1}{\epsilon}\right)\theta}{1 - \dfrac{\theta}{\rho_1}},$$

$$\mu_2 \text{ (tax on saving)} = \frac{1}{\epsilon}\theta,$$

and we see that income-tax should be partially but not wholly remitted on savings. The case for remission would, however,

[1] No account is taken of graduation in this.

[2] Strictly, we consider the limit as this time tends to zero.

be strengthened enormously by taking into account the expectation of taxation in the future.

(23) It should be emphasized in conclusion that the results about " infinitesimal " taxes can only claim to be approximately true for small taxes, how small depending on data which are not obtainable. It is perfectly possible that a tax of 500% on whisky could for the present purpose be regarded as small. The unknown factors are the curvatures of the supply and demand curves; if these are zero our results will be true for any revenue whatever, but the greater the curvatures the narrower the range of "small" taxes.

On the other hand, the more complicated results contained in equations (3), (3′), (11), (13) may well be valid under still wider conditions. But these are, in the general case, too complicated to be worth setting down in the absence of practical data to compare with them.

APPENDIX

We can also say something about the more general problem in which the State wishes to raise a revenue for two purposes; first, as before, a fixed money revenue, R_1, which is transferred to rentiers or otherwise without effect on the demand schedules; and secondly, an additional revenue, R_2, sufficient to purchase fixed quantities, $a_1, a_2, \ldots a_n$ of each commodity.

Let us denote by p_r, q_r, as before, the demand and supply prices of the rth commodity, and the tax on it by λ_r. Then if x_r is the amount of the rth commodity consumed by the public (or by the State out of R_1), $x_r + a_r$ is the amount produced, and we have

$$\frac{\partial u}{\partial x_r} = \lambda_r = p_r(x_1, x_2, \ldots, x_n) - q_r(x_1 + a_1, x_2 + a_2, \ldots, x_n + a_n),$$

$$R_1 + R_2 = \Sigma \lambda_r x_r, \quad R_2 = \Sigma a_r q_r;$$

so that u is to be a maximum subject to

$$\Sigma \lambda_r x_r - \Sigma a_r q_r = R_1 = \text{constant},$$

whence

$$\frac{\lambda_r}{\Sigma_s x_s \dfrac{\partial \lambda_s}{\partial x_r} - \Sigma_s a_s \dfrac{\partial q_s}{\partial x_r}} = \quad = \quad = -\theta$$

or

$$\frac{\lambda_r}{\Sigma_s (a_s + x_s) \dfrac{\partial q_s}{\partial x_r} - \Sigma_s x_s \dfrac{\partial p_s}{\partial x_r}} = \theta, \text{ which replace equations (3).}$$

Although these equations do not give such simple results as we previously obtained for an infinitesimal revenue or a quadratic utility function, in the cases considered in § 15 and § 17 they lead us again to the equations (11) and (13).

For, taking the case of § 15, in which the commodities are independent both for demand and supply, and, as before, denoting by μ_r the rate of tax *ad valorem* on the rth commodity and by ρ_r, ϵ_r its elasticities of demand and supply for the amounts x_r, $x_r + a_r$ respectively consumed and produced by the public, we have

$$\frac{\mu_r}{\dfrac{x_r + a_r}{q_r}\dfrac{dq_r}{d(x_r + a_r)} - \dfrac{x_r}{q_r}\dfrac{dp_r}{dx_r}} = \theta$$

or

$$\frac{\mu_r}{\dfrac{1}{\epsilon_r} + \dfrac{1 + \mu_r}{\rho_r}} = \theta$$

whence $\mu_r = \dfrac{\left(\dfrac{1}{\epsilon_r} + \dfrac{1}{\rho_r}\right)\theta}{1 - \dfrac{\theta}{\rho_r}}$, which is equation (11) again. And we

can similarly derive equation (13) from the assumption of independence for demand and equivalence for supply.

<div style="text-align: right">F. P. RAMSEY</div>