NBER WORKING PAPER SERIES

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

Joseph E. Stiglitz

ABSTRACT

Frank Ramsey's classic paper "A contribution to the theory of taxation" gave rise to the modern theory of optimal taxation. This paper traces the literature that grew out of Ramsey's 1927 paper and assesses which of its key insights has proven robust. Though the path breaking work of Peter Diamond and James Mirrlees showed that Ramsey's results could be generalized 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 recognizing 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.

Joseph E. Stiglitz
Uris Hall, Columbia University
3022 Broadway, Room 212
New York, NY 10027
and NBER
jes322@columbia.edu

In Praise of Frank Ramsey's Contribution to the Theory of Taxation[1]

Joseph E. Stiglitz[2]

Frank Ramsey's brilliant 1927 paper, modestly entitled, "A contribution to the theory of taxation," is a landmark in the economics of public finance. A third of a 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 revolutionizing public finance.[3]

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 (1926)).[4] 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 maximizes 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 behavior, 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 labor 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[5].)

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

[2] University Professor, Columbia University
[3] Given the huge number of papers that have been written in this field in the last half century, this paper cannot even attempt a survey. Rather, focus on the central themes raised by Ramsey in his paper.
[4] In addition, he made important contributions to mathematics, logic, and philosophy, before dying at the young age of 26.
[5] See the discussion below and the references cited there.

1

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.") Similar reasoning would suggest that if there are a variety of tools for managing the macro-economy, in general, they should all be employed.

In the first section of this short celebratory article, I describe briefly Ramsey's basic insights (section i), and the early history of the development of the ideas based on Ramsey's paper (section ii). 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 not only does he provide a general formulation of the set of optimal taxes, but he 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 elasticity of demand and supply.

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

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 labor is supplied inelastically, then optimal taxation requires imposing a tax on labor alone ; but a proportionate tax on all commodities is equivalent to a tax on labor,. This is hardly a surprising result—one for which one didn't need Ramsey's apparatus (as Ramsey himself recognized): When there is an inelastic factor, it has long been recognized that there is no distortion associated with taxing it.[6]

---

[6] Henry George (1886) focused on the taxation of land as the inelastically supply factor of production. Subsequent research by Vickrey (1977, 1999) and Arnott and Stiglitz (1978) 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.)

2

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, Lipsey and Lancaster (1956) and Meade (1955) 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 characterize optimal resource allocations under such conditions. But Ramsey had shown 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) 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 e.g. to the theory of capital taxation; extending the results, e.g. to more explicitly general equilibrium models); and to understanding the limits of the results. This paper will briefly note some key examples in each of these areas.

2. *Earlier dissemination*

While Ramsey's work was incorporated into Pigou's classical public finance textbook (1928), 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 realizing 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)[7], in the context of regulatory policy—analyzing 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 maximizing government would impose similar prices suggested that perhaps monopoly was not as bad as had previously been thought; and that conclusion became reinforced in those contexts where monopoly profits are driven to zero by

---

[7] English translation 1971. See also Baumol and Bradford, 1970.

contestability[8] (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" maximizing 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 Peter Diamond and James 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.)[9]

 Though the most general Diamond-Mirrlees' result on equi-proprtionate reductions in output of each commodity along the compensated demand curves[10] seemed fully parallel to that of Ramsey, the exploration of specialized 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, when one cannot impose lump

---

[8] 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.

[9] As we note below, this result was not general.

[10] As Dasgupta and Stiglitz (1971) point out, what is required is an equi-proportionate reduction of 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.

4

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 hypothesized), 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 Collett and Hague: (a)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; (b) especially increase taxes on products that provide an indirect tax on rents/profits; and (c) 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.[11] 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.[12]

  Some examples illustrate: Consider gasoline.   Gasoline used by consumers can be considered an input into the production of transportation services—that is the final consumption good being produced. Production efficiency requires not taxing an input, but taxing the output (transportation services), but that output cannot be taxed (that is, one cannot observe directly the amount the car is driven)[13]. Hence, it is "optimal" to tax the input of gasoline as a *surrogate* for the (missing) tax on the output, even though 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 (gasoline.) On the other hand, if we do tax the pure profits (rents) in the oil industry, we need

[11] This result was very much in the spirit of Ramsey and Diamond-Mirrlees, who emphasized the effect of an increase in the tax on one commodity on the revenues raised by other taxes.

[12] At the same time, they showed that when the rents of different producers can be taxed at different rates, production efficiency is still desirable.

[13] 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 don't 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.

5

to adjust the optimal tax on gasoline 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, fertilizer, 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. (See Emran and Stiglitz (2005) and 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)[14]:

(2)   t/p = 1/elasticity of demand + (1- τ)/elasticity of supply

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 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 emphasized 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; that is because it is not possible for government to observe most transactions, e.g. among farmers. (See Sah and Stiglitz, 1992). By the same token, we know that some of the returns to a firm in excess of payments to labor 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. (See 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. 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 for the untaxed good, and it may be desirable to use this surrogate. Alternatively, an increase in taxes on the freely taxed goods may

---

[14] These are compensated elasticities.

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.[15] 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, e.g. 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 didn't 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 explicitly focus 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, e.g. by Atkinson and Stiglitz (1972, 1980), and depend on the weight given to equity.[16])

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

---

[15] 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.
[16] See also Diamond, 1975; Mirrlees, 1975.

7

always necessary to be precise about what are the constraints: even if lump sum taxes are not feasible, other taxes may be. Most importantly, the government can impose redistributive income taxes, or even possibly redistributive wage taxes.

Once one goes 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.[17] 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).)

The case of an optimal (non-linear) income tax is very special. Countries at best approximate the optimal tax system with a piece-wise linear tax structure[18]. Though Kaplow(2006) argues that the Atkinson-Stiglitz results do not depend on the income tax being imposed being optimal, as in the general theory of third and fourth best taxation, there are still circumstances in which there may be some role for redistributive commodity taxation.

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[19]. 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.[20]

6. Non-linear taxation and heterogeneous individuals

---

[17] The concept of Pareto efficient taxation is noted in Mirrlees (1976) and Stiglitz [1982], and explored further in Stiglitz [1985, 1987, 1998] and Brito *et al* 1990
[18] See Stiglitz (2009).
[19] For a more general discussion of Pigouvian corrective taxes, see the commentary below.
[20] This is the essential point of Greenwald and Stiglitz (1986) and Arnott, Greenwald, and Stiglitz (1994)

Moreover, once one admits the possibility of non-linear *income* taxes, shouldn't 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 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, e.g. 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 percent, 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* behavior, e.g. in the context of risk taking) is by 1 to 2%. But if individuals have different indifference curve (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.[21]

---

[21] 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.

9

7. An information theoretic perspective

There are many ways that individuals differ that are easily observable, e.g. age, gender, marital status, and some of these (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 behavior, 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.[22] Thus, with separability of utility (between consumption and leisure) and the only basis of individual differences being labor 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.[23] 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 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 subsidize 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 criticized these subsidies, the optimal tax theory growing out of Ramsey's seminal work provides a rationale.[24]

8. General Equilibrium incidence

We have described how, at the center of modern optimal tax theory, and the work growing out of Ramsey, is 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 labor 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, and in ways that can reduce the burden imposed on distortionary redistribution policies. If it is possible to do so, governments should--again, even possibly if it interferes with production efficiency, and even if in doing so, it seemingly reduces the progressivity of the tax system.

---

[22] 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.

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

[24] 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.

Thus, assume there are two kinds of labor, skilled and unskilled, which are complements. By encouraging greater supply of skilled labor, the relative wages of the unskilled labor will increase. If skilled laborers 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.[25]

9. Interest income tax

Ramsey's analysis also had important and profound implications concerning the taxation of interest income, which he himself recognized. We can treat current and future consumption as two different commodities.[26] 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 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 infinite lived individuals with constant utility discount rate model, the long run supply elasticity of capital is infinite, which means that there should be no taxation of capital.[27]

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.[28]

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 from their parents large amounts of human or financial capital. Leveling the playing field--providing more equality of opportunity--necessitates a tax on such cross generational distributions. We cannot, however, monitor such transfers perfectly, but

---

[25] 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 labor. 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 Mieskowsi (1969). In the context of the subject at hand, see Diamond (1978).
[26] 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.
[27] Judd (1985) , Stiglitz (1985) and Chamley (1986). As Saez and Pikkety (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.
[28] See, for instance, Mankiw *et al* (2009), Atkeson *et al.* (1999), Lucas (1990).

capital income may be correlated with such transfers (especially at the upper end), and if so, they provide an appropriate basis for taxation. [29]

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)[30])

General equilibrium effects provide a more ambiguous basis of capital taxation. If capital is a substitute for unskilled labor (as seems increasingly the case with robotization) but a complement of skilled labor (an implication of skill-biased innovation) then a capital tax might not only improve the after tax distribution of income but the before tax distribution of income. There can also be distributional effects through the demand side, e.g. taxation of capital intensive consumption goods and subsidizing unskilled labor intensive goods may lead to an improvement in the before-tax-and-transfer distribution of income. [31]

One of the most important reasons for taxing capital income is that we cannot clearly distinguish capital income from wage income, particularly the labor 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 labor? (This question has significant policy resonance: in the US, private equity managers have attempted to receive capital gains treatment for

---

[29] A variety of papers have explored these ideas. See in particular Cremer, Pestieau, and Rochet (2003), Farhi and Werning (2010), and Piketty and Saez (2012). As Piketty and Saez 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 analyzing bequest behavior 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.

[30] 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.

[31] Stiglitz (1985, 1998) emphasizes, 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 generation 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.

12

their income, which in most cases is simply a return to their labor. More generally, when we tax capital income preferentially, there are efforts to convert wage income into capital income.)[32] [33]

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 labor. Ramsey, and most of the subsequent literature, *assumed* that markets were competitive, that the economy was well-described by a neo-classical model. But today, it is increasingly being recognized that much of the income at the top is derived from rents (monopoly rents, from the ability of corporate CEO's 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 capitalized, 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.[34]

10. Ramsey taxation and Pigouvian corrective taxation

Ramsey essentially assumed a well-functioning market economy[35]. 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.[36]

Dasgupta and Stiglitz' (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-

---

[32] 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 labor 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 hypothesized, it might even be desirable to subsidize bequests (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.) (See Stiglitz, 1976, Bevan and Stiglitz, 1979, and Farhi and Werning 2010.)

[33] 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 (1982c).)

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

[35] 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.

[36] 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) and Greenwald and Stiglitz (2014) have derived Ramsey-like formulae for such corrective taxation.

13

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.[37]

Arnott and Stiglitz (1986) obtained formulae describing optimal taxes in the special case of distortions arising from moral hazard.[38] 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 behavior of consumers directly and of insurance companies; and insurance companies attempt to affect the behavior of those they insure. As they observe:

> "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."

Perhaps the most interesting and controversial question in this area is the validity of the "double dividend" associated with Pigouvian taxes--they raise revenues at the same time that they improve the efficiency of the economy. The question is to what extent do environmental taxes reduce (or augment) tax revenues raised from other sources, e.g. income taxation. Some (like Nordhaus, 1993) have argued that the revenue benefits of a carbon tax lead to a higher optimal level of the carbon tax, while many others argue that adverse effects on revenue suggest a lower tax rate.[39]

---

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

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

[39] For instance, Parry (1994) argues that the optimal tax is 63 to 78% of marginal environmental damages. Bovenberg and Goulder (1996) come to a similar conclusion. They argue further that in the presence of what they

14

While there are circumstances in which these adverse revenue effects might arise (for instance, if imposing a carbon tax led workers to work so much less that income tax revenues would diminish more than the direct revenue raised through the carbon tax), I believe that plausible restrictions on preferences ensure that there is still a double dividend. (See Stiglitz, (2013)) Indeed, advocates of a carbon tax suggest that imposing such a tax would induce firms to invest more, to retrofit their capital stock to reflect the new set of factor prices, and especially in a world in which there are demand constraints on output, this could substantially increase revenues. In any case, the presumption in favor of corrective taxation (moderated by these potential adverse revenue effects) remains.

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 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 recognize 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.[40]

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

---

view as realistic policy constraints the optimal carbon tax rates are "far below the marginal environmental damages--and may even be negative." (p. 994)

[40] 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.

the imposition of progressive taxation more generally—is that such taxes may reduce the supply of capital or of skilled labor.[41]

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.[42]

12. Concluding Comments

Ramsey's paper can be viewed as the seminal paper in what has become an immense literature on optimal taxation.[43] He focused narrowly on the question of how to minimize 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.[44]

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 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 generalized 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 recognizing 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.

---

[41] 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. (See, e.g. Stiglitz 1977, 1983a, 1983b). This analysis has an obvious and immediate application to attempts to impose redistributive taxes within the EU.

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

[43] 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 (see, e.g. Stiglitz (1998)) . This is a sub-theme within the huge literature in cost benefit analysis.

[44] 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. Arnott and Stiglitz (1988), Brito *et al* 1995 , and Stiglitz (1983b)

16

Ramsey ended his 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." 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 minimizing the deadweight loss of our tax system, we would have imposed lump sum taxes: Tax structures that make use of observable information that enable us to more closely approximate the kind of taxation we would have imposed 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.)

References

Aiyagari, S. Rao, 1995, "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy* 103(6): 1158-75.

Arnott, R. and J. E. Stiglitz, 1979, "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size," *Quarterly Journal of Economics*, 93(4): 471-500.

—— and , 1986, "Moral Hazard and Optimal Commodity Taxation," *Journal of Public Economics*, 29: 1-24.

_____ "Randomization with Asymmetric Information," *Rand Journal of Economics*, 19(3), Autumn 1988, pp. 344-362

Arnott, R. and B. Greenwald and J. E. Stiglitz, 1994 "Information and Economic Efficiency," *Information Economics and Policy,* 6(1): 77-88.

Atkeson, A., V. V. Chari, and P. J. Kehoe, 1999, "Taxing Capital Income: A Bad Idea," *Federal Reserve Bank of Minneapolis Quarterly Review,* 23(3): pp. 3-17.

Atkinson, A., 1980. "Inheritance and the Redistribution of Wealth", in G.M. Heal and G.A. Hughes, eds., "Public Policy and the Tax System," London: Allen and Unwin, pp. 26-66.

Atkinson, A. B. and J. E. Stiglitz, 1972, "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics*, 1(1): 97-119

—— and ——, 1976, "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, 6(1): 55-75.

—— and ——, 1980, *Lectures on Public Economics*, New York and London: McGraw-Hill Book Company.

Baumol, W. J. and D. F Bradford, 1970, "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, 60(3): 265-83.

William J. Baumol, John C. Panzar, & Robert D. Willig (1982). Contestable Markets and the Theory of Industry Structure: San Diego, CA : Harcourt Brace Jovanovich.

Bevan, D. and J. E. Stiglitz, 1979, "Intergenerational Transfers and Inequality," *The Greek Economic Review*, 1(1): 8-26.

Boadway, R. and M. Sato, 2009 "Optimal Tax Design and Enforcement with an Informal Sector," *American Economic Journal: Economic Policy* 1(1): 1-27

Boiteux, M., 1956, "Sur la gestion des monopoles publics astreints a l'equilibre budgetaire," *Econometrica*, 24(1): 22-40.

——, 1971, "On the management of public monopolies subject to budgetary constraints," *Journal of Economic Theory*, 3(3): 219-240.

Bovenberg, A. Larns and L. Goulder, 1996, "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses," *American Economic Review,* v. 86, no. 4 (sept), pp. 985-1000.

Brito , D.L., J.H. Hamilton S.M. Slutsky and J. E. Stiglitz, 1990,"Pareto Efficient Tax Structures," *Oxford Economic Papers*, 42, pp. 61-77.

——, ——, ——, and ——, 1995, "Randomization in Optimal Income Tax Schedules," , *Journal of Public Economics*, 56(2): 189-223.

William A. Brock, 1983, "Contestable Markets and the Theory of Industry Structure: A Review Article," *The Journal of Political Economy*, 91(6): 1055–1066.

Chamley, C., 1986, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica* 54(3): 607-622.

——, 2001, "Capital Income Taxation, Wealth Distribution and Borrowing Constraints," *Journal of Public Economics* 79(1): 55-69.

Corlett, W.J. and D.C. Hague, 1953, "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies* 21(1): 21-30.

Conesa, J.C., S. Kitao, and D. Krueger, 2009, "Taxing Capital? Not a Bad Idea after All!" *American Economic Review* 99(1): 25-48.

Cremer, H., F. Gahvari, and N. Ladoux, 1998 "Externalities and optimal taxation," *Journal of Public Economics,* 70, pp. 343-364

Cremer, H. and P. Pestieau, 2006, "Wealth Transfer Taxation: A Survey of the Theoretical Literature," in S.C. Kolm and J. Mercier Ythier, eds, *Handbook of the Economics ofGiving, Altruism, and Reciprocity* 2(2): 1107-1134.

Cremer, H., P. Pestieau, and J.-C. Rochet, 2003, "Capital Income Taxation When Inherited Wealth Is Not Observable," *Journal of Public Economics* 87(11): 2475-90.

Dasgupta, P. and J. E. Stiglitz, 1971, "Differential Taxation, Public Goods, and Economic Efficiency," *Review of Economic Studies* 38(2): 151-174.

—— and ——, 1972, "On Optimal Taxation and Public Production," *Review of Economic Studies* 39(1): 87-103.

De Bonis, V. and L. Spataro,2005, "Taxing Capital Income as Pigouvian Correction: The Role of Discounting the Future," *Macroeconomic Dynamics,* 9, 469-477.

Diamond, P., 1975, "A Many-Person Ramsey Tax Rule," *Journal of Public Economics* 4(4): 335-42.

——, 1978, "Tax Incidence in a Two Good Model," *Journal of Public Economics*, 9(3): 283-299.

—— and J.A. Mirrlees, 1971, "Optimal Taxation and Public Production," *American Economic Review* 61(1):8-27 and 261-278.

Emran, S. and J. E. Stiglitz, 2005, "On Selective Indirect Tax Reform in Developing Countries," *Journal of Public Economics* 89(4): 599-623.

Farhi, E. and I. Werning, 2010, "Progressive Estate Taxation," *Quarterly Journal of Economics* 125(2): 635-673.

Farrell, J., 1986, "How Effective is Potential Competition?" *Economics Letters* 20(1): 67-70.

Greenwald, B. and J. E. Stiglitz, 1986, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics* 101(2): 229-264.

—— and ——, 2014, "Learning, Growth, and Development: A Lecture in Honor of Sir Partha Dasgupta," in *Environment & Development Economics: Essays in Honor of Sir Partha Dasgupta,* Scott Barret, Karl-Goran Maler, and Eric S. Maskin (eds.), Oxford: Oxford University Press, pp. 31-94

Harberger, AC., 1962, "The Incidence of the Corporate Income Tax," *Journal of Political Economy* 70(3): 215-240.

Judd, K., 1985, "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28(1): 59-83.

——, 1999, "Optimal taxation and spending in general competitive growth models," *Journal of Public Economics*, 71(1): 1-26.

Kaplow, L., 2006, "On the Undesirability of Commodity Taxation Even When Income Taxation Is Not Optimal," *Journal of Public Economics* 90(6/7): 1235-1250.

Keen, M., 2008, "VAT, Tariffs, and Withholding: Border Taxes and Informality in Developing Countries," *Journal of Public Economics*, 92(10): 1892-1906.

Konishi, H., 1995, "A Pareto-Improving Commodity Tax Reform Under a Smooth Nonlinear Income Tax," *Journal of Public Economics* 56(3): 413-446.

Laroque, G., 2005, "Indirect Taxation Is Harmful under Separability and Taste Homogeneity: A Simple Proof," *Economics Letters* 87: 141-144.

Lucas, R., 1990, "Supply-Side Economics: An Analytical Review," Oxford Economic Papers, New Series, 42(2): 293-316.

Mankiw, N.G., M. Weinzierl, D. Yagan, 2009, "Optimal Taxation in Theory and Practice," NBER Working Paper No. 15071, June 2009.

Mayeres, I. and S. Proost, 1997, "Optimal Tax and Public Investment Rules for Congestion Types of Externalities," *Scandinavian Journal of Economics,* v. 99, no 2, June, pp. 261-279.

Mieszkowski, P., 1969, "Tax Incidence Theory," *Journal of Economic Literature* 7(4): 1103-l 124.

Mirrlees, J.A., 1971, "An Exploration in the Theory of Optimal Income Caxation," *Review of Economic Studies* 38(2): 175-208.

——, 1972, "On Producer Taxation," *Review of Economic Studies* 39(1):105-111.

——, 1975, "Optimal Commodity Taxation in a Two-Class Economy," *Journal of Public Economics* 4(1): 27-33.

Nordhaus, W. D., 1993, "Optimal greenhouse gas reductions and tax policy in the "DICE" model," *American Economic Review,* 83, pp. 313-317.

Parry, I., 1994, "Pollution Taxes and Revenue Recycling," *Journal of Environmental Economics and Management,* 29, S-64-77.

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

Piketty, T., Emmanuel Saez, Stefanie Stantcheva, 2011, "Taxing the 1%: Why the Top Tax Rrate Could Be Over 80%," *Vox*, December 8.

Piketty, T. and E. Saez, 2012, "A Theory of Optimal Capital Taxation," NBER Working Paper 17989.

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

Ramsey, F.P., 1926, "Truth and Probability," Published 1931 in *Foundations of Mathematics and other Logical Essays*, Ch. VII, p.156-198. Edited by R.B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co. Ltd. New York: Harcourt, Brace and Company

——, 1927, "A contribution to the theory of Taxation," *Economic Journal,* 37(145): 47-61

——, 1928, "A Mathematical Theory of Saving," *Economic Journal* 38(152): 543-559

Saez, E., 2001, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies* 68(1): 205-229.

——, 2002, "The Desirability of Commodity Taxation under Nonlinear Income Taxation and Heterogeneous Tastes," *Journal of Public Economics* 83: 217-230.

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

Salanié, B., 2003, *The Economics of Taxation*, Cambridge, Mass.: MIT Press.

Sandmo, A, 1975, "Optimal Tazation in the Presence of Externalities," *Swedish Journal of Economics,* 77, no 1, March, pp. 86-98

——, 1976, "Direct vs. Indirect Pigovian Taxation," *European Economic Review,* 7, pp. 337-349.

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

—— and ——, 1987b, "Privatization, Information and Incentives," with D. Sappington, *Journal of Policy Analysis and Management*, 6(4): 567-582

Slemrod, J., 1990, "Optimal Taxation and Optimal Tax Systems," *Journal of Economic Perspectives* 4(1): 157-178.

Stiglitz, J. E., 1969, "The Effects of Income, Wealth and Capital Gains Taxation on Risk-Taking," *Quarterly Journal of Economics*, 83(2): 263-283.

——, 1972, "Taxation, Risk Taking, and the Allocation of Investment in a Competitive Economy," in M. Jensen, ed., *Studies in the Theory of Capital Markets*, New York: Praeger, pp. 294-361. (Proceedings of a conference at the University of Rochester, August 1969.)

——, 1976, "Notes on Estate Taxes, Redistribution, and the Concept of Balanced Growth Path Incidence," *Journal of Political Economy*, 86(2), part 2: 137-150.

——, 1977, "Theory of Local Public Goods," in M.S. Feldstein and R.P. Inman, eds., *The Economics of Public Service*, MacMillan Publishing Company, pp. 274-333. (Paper presented to IEA Conference, Turin, 1974.)

——, 1978, "Equality, Taxation and Inheritance," in *Personal Income Distribution*, W. Krelle and A.F. Shorrocks (eds.), North-Holland Publishing Company, 1978, pp. 271-303. (Proceedings of IEA Conference, Noordwijk aan Zee, Netherlands, April 1977.)

——, 1982a, "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics* 17(2): pp. 213-240.

——, 1982b "Utilitarianism and Horizontal Equity: The Case for Random Taxation," *Journal of Public Economics*, 18, 1982, pp. 1-33

——, 1982c, The Rate of Discount for Cost-Benefit Analysis and the Theory of the Second Best," in *Discounting for Time and Risk in Energy Policy*, R. Lind (ed.), Resources for the Future, 1982, pp. 151-204.

——, 1983a, "Public Goods in Open Economies with Heterogeneous Individuals," in J.F. Thisse and H.G. Zoller, eds., *Locational Analysis of Public Facilities,* Amsterdam: North-Holland, 1983, pp. 55-78.

——, 1983b, "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*, New York: Academic Press, pp. 17-53.

——, 1985, "Inequality and Capital Taxation," IMSSS Technical Report 457, Stanford University, July.

——, 1987, "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," in A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics, 2,* Amsterdam: North Holland, pp. 991-1042.

——, 1988, "Technological Change, Sunk Costs, and Competition," in M.N. Baily and C. Winston, eds, *Special Issue on Macroeconomics*, The Brooking Institution, pp.883-947.

——, 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)).

——, 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,* Oxford, UK: Oxford University Press, pp. 535-566. (revision of 1976 paper)

——, 2010, "Development-Oriented Tax Policy," in R. H. Gordon, ed., *Taxation in Developing Countries*, New York: Columbia University Press, pp. 11-36.

——, 2012, *The Price of Inequality: How Today's Divided Society Endangers Our Future*, New York: W.W. Norton.

——, 2013, Sharing the Burden of Saving the Planet: Global Social Justice for Sustainable Development," in Mary Kaldor and Joseph E. Stiglitz, eds., *The Quest for Security: Protection without Protectionism and the Challenge of Global Governance*, New York: Columbia Unversity Press, pp. 161-190

——, 2014, "Learning, Growth, and Development: A Lecture in Honor of Sir Partha Dasgupta," in *Environment & Development Economics: Essays in Honor of Sir Partha Dasgupta,* Scott Barret, Karl-Goran Maler, and Eric S. Maskin (eds.), Oxford: Oxford University Press, pp. 31-94

Vickrey, W., 1977, "The City as a Firm," in M. S. Feldstein and R. P. Inman, eds., *The Economics of Public Services,* London: Macmillan, 1977.

——, 1999, "Henry George, Economies of Scale, and Land Taxation," in K. C. Wenzer, ed. *Land Value Taxation*, New York: M. E. Sharpe.