WHAT HAS THE OPTIMAL TAXATION PARADIGM TAUGHT US?

Guillaume Allègre

SCIENCES PO OFCE WORKING PAPER n° 14/2025

EDITORIAL BOARD

CONTACT US

OFCE
10 place de Catalogne | 75014 Paris | France
Tél. +33 1 44 18 54 24
www.ofce.fr

WORKING PAPER CITATION

This Working Paper:
Guillaume Allègre, OFCE-Sciences Po,
What has the optimal taxation paradigm taught us?
Sciences Po OFCE Working Paper, n° 14/2025.
Downloaded from URL: www.ofce.sciences-po.fr/pdf/dtravail/WP2025-14.pdf
DOI - ISSN

ABOUT THE AUTHORS

Guillaume Allègre, OFCE-Sciences Po.
Email Address: guillaume.allegre@sciencespo.fr

RESUME

L'article de Mirrlees de 1971, An Exploration in the Theory of Optimum Income Taxation, est largement reconnu comme ayant façonné la théorie moderne de l'imposition sur le revenu. Mirrlees cherche à identifier les principes qui devraient guider la fiscalité des revenus en combinant des objectifs d'équité et d'efficacité. Le résultat de l'article est un barème fiscal optimal. Ce document de travail remet en question la pertinence du paradigme de la fiscalité optimale en revenant à ses origines dans l'article de Mirrlees. Le modèle de Mirrlees est normatif ; il discute de ce que la société devrait faire. Ce document soutient que le paradigme n'est pas convaincant dans la mesure où le lecteur n'est pas susceptible de modifier ses opinions sur les politiques en suivant le raisonnement. Le modèle est hypothético-déductif, mais les gens ont généralement des idées plus arrêtées sur les conclusions du modèle, en matière de taux marginaux d'imposition, que sur ses hypothèses, notamment le paramètre de préférence sociale pour l'égalité. De plus, un modèle normatif devrait être jugé selon sa pertinence normative, or les implications normatives du modèle sont rarement discutées. Les conditions de succès du paradigme n'ont jamais été établies. La promesse initiale de Mirrlees était de distinguer clairement les désaccords normatifs des désaccords descriptifs. Cette promesse n'a jamais été tenue.

MOTS CLES

Taxation optimale, Mirrlees, économie normative.

ABSTRACT

Mirrlees' 1971 article, An exploration in the Theory of Optimum Income Taxation is widely recognized as having shaped modern income tax theory. Mirrlees aims to find the principles that should govern income taxation by combining equity and efficiency objectives. The article's output is an optimal tax schedule. This working paper questions the relevance of the optimal taxation paradigm by returning to its origins in Mirrlees' article. Mirrlees' model is normative; it discusses what society should do. This paper argues that the paradigm is not convincing: people are not likely to change their views on policies based on this line of rea-soning. The model is hypothetico-deductive, but people usually have stronger ideas about the conclusions of the model in terms of marginal tax rates than about its hypothesis, in terms of a parameter for social preference for equality. Also, a normative model should be judged by its normative relevance, but the normative implications of the model itself are rarely dis-cussed. The paradigm's success conditions have never been examined. The initial promise of Mirrlees was to disentangle normative and descriptive disagreements. It was never fulfilled.

KEYWORDS

Optimal taxation, Mirrlees, normative economics.

JEL

H21, I38.

What has the optimal taxation paradigm taught us?

Guillaume Allègre OFCE – Sciences Po

Abstract. Mirrlees' 1971 article*, An exploration in the Theory of Optimum Income Taxation* is widely recognized as having shaped modern income tax theory. Mirrlees aims to find the principles that should govern income taxation by combining equity and efficiency objectives. The article's output is an optimal tax schedule. This working paper questions the relevance of the optimal taxation paradigm by returning to its origins in Mirrlees' article. Mirrlees' model is normative; it discusses what society should do. This paper argues that the paradigm is not convincing: people are not likely to change their views on policies based on this line of reasoning. The model is hypothetico-deductive, but people usually have stronger ideas about the conclusions of the model in terms of marginal tax rates than about its hypothesis, in terms of a parameter for social preference for equality. Also, a normative model should be judged by its normative relevance, but the normative implications of the model itself are rarely discussed. The paradigm's success conditions have never been examined. The initial promise of Mirrlees was to disentangle normative and descriptive disagreements. It was never fulfilled.

Keywords: optimal taxation, Mirrlees, normative economics.

JEL: H21, I38.

Que nous a appris le paradigme de fiscalité optimale ?

Résumé. L'article de Mirrlees de 1971, An Exploration in the Theory of Optimum Income Taxation, est largement reconnu comme ayant façonné la théorie moderne de l'imposition sur le revenu. Mirrlees cherche à identifier les principes qui devraient guider la fiscalité des revenus en combinant des objectifs d'équité et d'efficacité. Le résultat de l'article est un barème fiscal optimal. Ce document de travail remet en question la pertinence du paradigme de la fiscalité optimale en revenant à ses origines dans l'article de Mirrlees. Le modèle de Mirrlees est normatif ; il discute de ce que la société devrait faire. Ce document soutient que le paradigme n'est pas convaincant dans la mesure où le lecteur n'est pas susceptible de modifier ses opinions sur les politiques en suivant le raisonnement. Le modèle est hypothético-déductif, mais les gens ont généralement des idées plus arrêtées sur les conclusions du modèle, en matière de taux marginaux d'imposition, que sur ses hypothèses, notamment le paramètre de préférence sociale pour l'égalité. De plus, un modèle normatif devrait être jugé selon sa pertinence normative, or les implications normatives du modèle sont rarement discutées. Les conditions de succès du paradigme n'ont jamais été établies. La promesse initiale de Mirrlees était de distinguer clairement les désaccords normatifs des désaccords descriptifs. Cette promesse n'a jamais été tenue.

Mots clés : Taxation optimale, Mirrlees, économie normative.

Table of contents

1. A very brief history of optimal taxation until 1971 5
2. Mirrlees' model of optimal taxation (1971) 8
   2.1. A normative model 8
   2.2. How does Mirrlees' model work? 8
   2.3. The initial conclusions 11
   2.4. Extensions 12
3. What have economists learned from this model and its extensions? 13
   3.1 The greater the inequality, the greater the redistribution must be 13
   3.2 What is the optimal guaranteed minimum income? 14
   3.3 An optimal U-shaped or linear curve? 17
   3.4 Tagging (discrimination) is always optimal 19
4 Why the optimal taxation paradigm is unconvincing 21
   Hypotheses that are weaker than conclusions 22
   A model that must be judged on its relevance, but does not discuss it 22
   A model that is not based on shared intuitions 24
   From transparency to the argument from authority 24

Conclusion: opening the black box 25

What has the optimal taxation paradigm taught us?

Mirrlees' article*, An exploration in the Theory of Optimum Income Taxation*, published in 1971, is widely recognized as having "shaped modern income tax theory" and marked "the birth of modern incentive theory" (Simula and Trannoy, 2007). According to Piketty and Saez (2013), "The modern analysis of optimal income taxation started with Mirrlees (1971), who rigorously posed and solved the problem." This program was pursued in particular by Atkinson, Stiglitz, Diamond, Piketty, Saez, and others. The *Mirrlees Review* published by the Institute of Fiscal Studies (2010, 2011) is a reference work, and the paradigm of optimal taxation as initiated in 1971 is still very much alive.

Mirrlees aims to find the principles that should govern income taxation by combining equity and efficiency objectives. The model thus makes it possible to quantify the trade-off between equity and efficiency and to make precise public policy recommendations in the form of an optimal transfer schedule. The term "optimal taxation" is misleading because the model focuses on "optimal transfers" or "optimal redistribution." Taxation in the 1971 article could therefore be negative and thus included the payment of a minimum income for those with no primary income. The conclusion takes the form of an optimal function giving a negative or positive optimal transfer for each level of earned income.

Compared to previous work, the article's contribution is to look at what happens at both the bottom and the top of the scale, in fact, across the entire income scale. This requires a consistent overall argument. Transfers are financed: reducing taxes at one income level requires increasing them at another. Lowering *marginal* tax rates at one point on the income scale, to provide an incentive, also has the effect, *ceteris paribus*, of reducing the *average* tax rate for all higher incomes, and therefore revenue. To keep revenue constant, it is then necessary to increase these marginal rates... The model allows for a comprehensive, rather than piecemeal, analysis of the vertical transfer system.

This working paper is not the first to discuss the contributions of the literature on "optimal redistribution." Among the most important articles are Stiglitz (1987), Slemrod (1990), Brewer, Saez, and Shepard (2010), Piketty and Saez (2013), Mankiw et al. (2009), to which Bourguignon (2001) should be added. Of all these articles, Bourguignon (2001) is the most critical of the relevance of an approach that he himself has applied extensively. It is no coincidence that the article is in French, and published as a response to another article (by Autume, 2001). Slemrod's article (1990) is also quite critical, but mainly raises important internal criticisms, calling for improvements to the paradigm rather than abandoning the approach altogether. Mankiw et al. do not appear to make any judgments about the relevance of the approach: they answer the questions "what does the theory say?" and "is the theory followed in practice?" The contributions of Stiglitz (1987) and Piketty and Saez (2013) are part of textbooks that aim to describe the theory and its extensions with precision. They include highly critical elements but preserve the paradigm, at least on the surface in the latter case. Brewer et al. contribute a chapter to the *Mirrlees Review*, led by the Institute of Fiscal Studies, which aims to examine contemporary tax policy claims considering optimal taxation theory, and is therefore the least critical. All these articles either defend the paradigm or, when they criticize it, ultimately save it (albeit sometimes with considerable reservations). This is not surprising. First, internal

criticisms are often the ones that historically leave the deepest mark when they are taken seriously. Second, as a normative model, the optimal taxation paradigm is not refutable. A normative model concludes in terms of what should be done. It cannot be refuted by comparing it to reality, to behavior, or to actual public policies that say nothing about the ideal. At worst, one can shrug one's shoulders and think that it is a somewhat peculiar way of looking at things. But the problem is more serious: even if it is not refutable, a paradigm should respond to legitimate objections made against it. However, the objections are numerous, long-standing, and... unanswered.

Most of the following insights come from at least one of the articles cited above. Taken together, they paint a much more negative picture of the practical relevance of the approach than any of these contributions individually. Perhaps we should take Stiglitz's (1987) remark seriously:

> *These considerations—as well as the practical considerations involved in deciding on the appropriate cardinalization within the utilitarian framework—make the utilitarian approach questionable as a guide to policy (p. 34).*

The conclusion seems decisive, stemming from a dual practical and theoretical problem... but this conclusion is not treated as such in the rest of the article.

Perhaps the most important—albeit ambiguous in its intent—critique of the optimal taxation approach comes from Mankiw and Weinzierl (2010). In their article, they suggest that if one follows the optimal taxation paradigm, income taxation should take into account the physical size of taxpayers. Their conclusion is as follows:

> *Our results, therefore, leave readers with a menu of conclusions. You must either advocate a tax on height, or you must reject, or at least significantly amend, the conventional utilitarian approach to optimal taxation. The choice is yours, but the choice cannot be avoided.*

It is unclear whether the authors are being purists or making an important external criticism. Mankiw et al. (2009) pointed out the same impasse:

> *Nevertheless, the theory behind tagging would suggest a much broader application. In particular, tax schedules ought to vary systematically, the theory tells us, with gender, height, skin color, physical attractiveness, health, parents' education, and so on. No modern tax system has such variation.*

This raises a question: why would a theory whose initial objective is to combine the goals of equity and efficiency only be consistent by sacrificing some form of equity (in this case, non-discrimination based on height and physical characteristics)? Are equity and efficiency commensurable without losing meaning? By defining fairness so narrowly (as simply reducing vertical inequality), Mankiw et al. end up asking why taxes do not optimally discriminate based on height, gender, skin color, physical attractiveness, health, and parents' education. It is difficult to decide whether this is a provocative reference or a reductio ad absurdum argument. In any case, Mankiw and his co-authors do not take the optimal taxation model seriously as a model that could produce relevant recommendations for public policy:

> *Recent work recommends capital taxation that is regressive in labor income changes, according to which capital income is taxed for those who earn surprisingly little and subsidized for those*

> *who earn surprisingly much. Few economists advising political candidates or elected government officials would have the temerity to advance these ideas in any practical discussion of tax policy.*
> *Why not? One possibility is that theory is right and that policymakers and the public are slow to appreciate certain valuable but counterintuitive insights. (p. 171-172).*

What is the point of a study that concludes with a tax recommendation that cannot be put forward in a practical discussion on tax policy? According to the authors, one possibility is that the theory is *correct* but misunderstood, but no criteria are put forward to define whether the theory is "*correct*" or not. What exactly does "*theory is correct*" mean in this context? The paradigm does not seem to provide any criteria for success. In this respect, the paradigm does not meet the criterion set by Blaug (1980):

> *What methodology can do is to provide criteria for the acceptance and rejection of research programs, setting standards that will help us to discriminate between wheat and chaff. (p.264)*

This working paper questions the relevance of the paradigm by returning to its origins in Mirrlees' article (1971).

1. A very brief history of optimal taxation until 1971

The question of taxation has preoccupied economists since the dawn of political economy. In *The Wealth of Nations* (1776), Smith offers four maxims concerning taxation. According to the first, "*the subjects of a state ought to contribute to the support of the government, each according to his ability, that is, according to his means, and not according to his wants.*" According to the second, "*the tax or portion of tax that each individual is required to pay must be certain and not arbitrary.*" Finally, according to the third, "*every tax should be paid at the time and in the manner which may be supposed to be least inconvenient to the taxpayer*." They are nothing more than what they claim to be, maxims, or brief statements of rules of good management.

Normative economics will continue to deliver such maxims or principles, but it will also take a more "scientific" turn by drawing on Benthamian utilitarianism. Utilitarianism has the advantage of being both a descriptive paradigm that assumes individuals maximize their well-being (assumption of rationality) and a normative doctrine that society should maximize a social welfare function, which is the pure aggregation of individual well-being. Positive and normative economics thus converge on the same paradigm, which is a gain in both consistency and aesthetics. Economists are very attached to the "elegance" of their models (McCloskey, 1983).

Normative utilitarianism benefits from a symmetry with the positive part of the discipline. This has the “charm of symmetry” and is also desirable because it allows economists to move from description to prescription in a coherent manner.

The utilitarian paradigm has other qualities that appeal to economists. The doctrine is individualistic (only individuals matter), consequentialist (only their well-being matters, not the procedure for achieving it), and impartial (all individuals are *a priori* equally valuable).

However, a schism in the discipline arose at the end of the 19th century with the marginalist revolution. Seligman (1908 [1894]) argued in several lengthy works that progressive income

taxation should be defended based on contributive capacity rather than diminishing marginal utility.

> *Are we then to abandon progressive taxation in theory? Before answering this question, it will be desirable to revert to the fundamental conception of faculty or ability, which is after all the best standard we have of the measure of general obligation to pay taxes, and to seek to ascertain what the faculty theory in its wisest interpretation can teach us in the matter. (p. 290)*

Edgeworth (1897) acknowledges that the argument is convincing... but does not belong to the realm of "pure theory":

> *The view thus distinguished is that according to which the sacrifice felt by the taxpayer is a dominant factor in the apportionment of the fiscal burden, the hedonistic, or in a special sense utilitarian, principle of taxation, as it may be called. Some other principle may be held—for instance, that of "ability," or "faculty," in a more objective sense—but can hardly be held to belong to the domain of pure theory.*
>
> *The purest, as being the most deductive form of utilitarianism, is that from which Bentham reasoned down to equality. There are those who regard this form as also purest, in that its first principle is the most apt to be universally accepted. That principle proposes as the end of action, or criterion of conduct, the greatest sum-total of happiness (p.550).*

Although he mentions "universal acceptance", Edgeworth is more interested in the elegance of the model: it is its purity that is supposed to make it acceptable. But Edgeworth does not respond to the criticisms of utilitarianism raised by Seligman (here in a later edition, 1908):

> *The imposition of "equal sacrifices" on all taxpayers must always remain an ideal impossible of actual realization. Sacrifice denotes something psychical, something psychological. A tax takes away commodities that are something material, something tangible. To ascertain the exact relations between something psychical and something material is impossible. No calculus of pain and pleasure can suffice, for no attempt to reduce the heterogeneous to the homogeneous can ever succeed.*

The sacrifice demanded by taxation cannot be measured, because this psychological quantity cannot be measured. Seligman's criticism extends to the entire Benthamian paradigm (which Bentham was aware of).

Although not universally accepted, Edgeworth's "total sum" has the merit of purity. It is necessary to have a pure economic theory of taxation, and utilitarianism provides this, allowing economic justification to be disconnected from other modes of justification. If social utility must be maximized, then Edgeworth shows that what society must equalize is not resources or rights but... individual marginal utilities[1] . However, economists will encounter the problem of inter-individual comparisons of utilities. This criticism is long-standing, having already been clearly formulated by Seligman, but the consensus in the 1930s was that such comparisons were impossible. According to Robbins (1932):

[1] In the 1970s, a moral philosophy debate asked the question, "Equality of What?" : we are all egalitarian, but what should be equalized? According to Sen (1979), the different answers are: marginal utility, total utility, primary goods, and capabilities.

> *But it is one thing to assume that scales can be drawn up showing the order in which an individual will prefer a series of alternatives, and to compare the arrangement of one such individual scale with another. It is quite a different thing to assume that behind such arrangements lie magnitudes which themselves can be compared as between individual scales. This is not an assumption which need anywhere be made in modern economic analysis, and it is an assumption which is of an entirely different kind from the assumption of individual scales of relative valuation (p. 122)*

More specifically:

> *The Law of Diminishing Marginal Utility does not justify the inference that transfers from the rich to the poor will increase total satisfaction. It does not tell us that a graduated income tax is less injurious to the social dividend than a non-graduated poll tax. Indeed, all that part of the theory of Public Finance which deals with "Social Utility" goes by the board. Interesting as a development of an ethical postulate, it is entirely foreign to the assumptions of scientific economics (p. 125).*

Bentham and Edgeworth's "greatest sum of happiness" cannot therefore be a universally accepted principle insofar as this criterion is not operational[2] .

The discipline then moved towards social choice theory, which consists of deriving public decisions from the rankings that individuals make of the different possible states of society. This is known as *New Welfare Economics*. Unfortunately, Arrow (1951) demonstrated that this type of hierarchy cannot respect several desirable basic properties. It was in this context that *the new new welfare* economics—as Stiglitz called it—developed in the 1970s and 1980s, notably with Atkinson, Diamond, Stiglitz, and, first and foremost, Mirrlees.

The way in which they resolve the problem of inter-individual comparison of utilities is summarized by Buchanan (1964) in a critique:

> *Economists, paying heed to Robbins, now know when they cross the bridge; they explicitly state their own value judgments in the form of "social welfare functions." Once having done this, they feel free to maximize to their own heart's content. And they do so within the bounds of methodological propriety, a la Robbins. They have, of course, abandoned his neutrality-of-ends position, but they have been straightforward about this.*

According to the new paradigm, utility maximization is therefore not THE “scientific” norm to follow, but one norm, the one followed by welfare economists. Other economists can follow other norms if they wish (and still be published), but once this norm is adopted, it is possible to derive normative conclusions in terms of fiscal policy. If economics is what economists do, then normative economics is welfare economics because it is what normative economists do.

The optimal taxation model is a good illustration of this.

[2] Hill (1968) identifies John Neville Keynes, Lionel Robbins, and Milton Friedman as the three leading figures of positivism in economics. The first is a cardinal utilitarian, the second rejects Benthamite utilitarianism, and the third rejects welfarism in general. The boundaries of the debates on utilitarianism and positivism therefore do not overlap.

2. Mirrlees' model of optimal taxation (1971)

2.1. A normative model

Mirrlees asks the question of the optimal income redistribution schedule. The question is: *what ought to be done?* What is the desirable marginal income tax rate? How much should be transferred to the poorest? What should the transfer schedule between these two points look like?

These are, therefore, questions of normative economics, which provide prescriptions in terms of public policy. Since the conclusion of Mirrlees' model is a normative conclusion, we can conclude that his model is normative, even though it also takes descriptive data as input.

2.2. How does Mirrlees' model work?

The model formalizes the trade-off that governments face when redistributing wealth from the rich to the poor. Edgeworth (1897) showed that utilitarianism implies a certain form of progressivity: if all individuals have the same utility function and this function implies a decreasing marginal utility of consumption, then redistributing from the wealthiest to the least wealthy increases social utility (equal to aggregate individual utilities). There is, therefore, a utilitarian gain from a progressive tax system and redistribution. Mirrlees supplements Edgeworth's intuition with behavioral responses. According to Mirrlees, individuals respond to redistribution by reducing their working hours and thus production. There is thus a trade-off between equality and efficiency in terms of production, size, and egalitarian distribution of the pie.

Formally, individuals have identical preferences and maximize the same utility function U(C, L), which increases with consumption and leisure, and therefore decreases with working time. Everyone chooses the amount of working time that maximizes their own utility. Individuals differ only in their productivity: more productive individuals obtain higher consumption per hour worked, with hourly wages assumed to be equal to their productivity. The government has at its disposal a unified socio-fiscal instrument with two functions: financing a public good and redistributing income under the constraint of financial equilibrium.
There is a solution to the model because individual utility functions are assumed to be concave: the marginal utility of consumption and leisure time is decreasing. The government thus faces a trade-off between equality and production: increasing taxes to finance transfers to the less well-off reduces inequality, which increases aggregate welfare, but reduces the supply of labor and total production (with taxation, work becomes less attractive because it allows for less consumption). Mirrlees demonstrates that, under several assumptions, it is possible to derive an optimal tax schedule.

The model is elegant. Social utility depends only on individual preferences, which are added together without weighting, so that everyone has equal weight and everyone's preferences also have equal weight. However, Mirrlees does not sweep social norms under the rug. He acknowledges that the existence of a social utility function that incorporates trade-offs between equality and efficiency inevitably implies the existence of social norms. These social norms are reflected in individual utility functions that are identical for everyone. Decreasing marginal utility—assumed to be equal for everyone—means that the degree of concavity of

this function is *de facto* a parameter of social preference for equality. The more concave the assumed unique function is, the greater the reduction in consumption inequalities increases social utility at equal total consumption.

Subsequently, this way of modeling social preferences was often abandoned—notably by Mirrless himself—so that the social preference parameter for equality would appear directly in the social utility function. This also makes it possible to distinguish this parameter from the form of individual utility functions (which is a parameter that attempts to describe behavior and is therefore positive). Aversion to inequality can be reflected, for example, by a different weighting in the social utility function of the consumption of the wealthiest individuals. Mirrlees (1976) demonstrates that there is no loss of generality in modeling in this way. The method is more transparent (and the calculation is easier to solve), but the initial liberal idea of assigning equal weight to everyone is lost.

It's sometimes said that Mirrlees' model is about limited information, not limited instruments. This is not true in the seminal 1971 article. According to the fifth assumption:

> *The State is supposed to have perfect information about the individuals in the economy, their utilities and, consequently, their actions.* (p. 176)

In theory, the State would like to tax the skill level of individuals. Mirrlees suggests that it is not fully practicable. This can mean other things than pure information constraints. Mirrlees mentions tax evasion in that sense. Skills are not fully measurable, but they are also endogenous[3]. Stiglitz (1982) transformed the problem into an asymmetric information argument. In this article, high-ability individuals can disguise themselves as low-ability individuals; optimal taxation should work as a screening device:

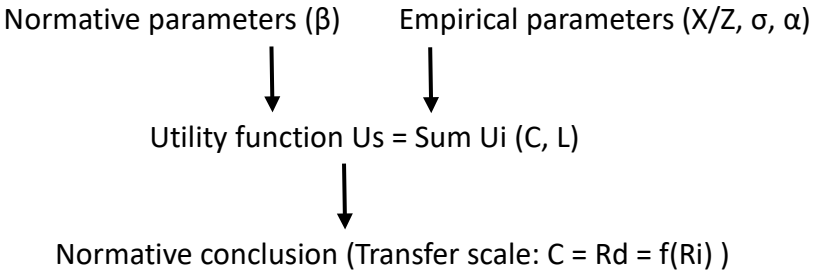
> *we make the lower ability individual's 'package' less attractive to him. (Since in this model both groups have identical utility functions, the only difference in the evaluation of a given consumption bundle arises from the differences in the leisure which they enjoy at any given level of income.) We thus can tax the higher ability individual more heavily without having him trying to 'disguise' himself as a low ability person. (p. 238)*

This formulation was popular in academic circles at the time and Mirrlees later used it. However, Mirrlees' initial argument is an incentive argument, a way to solve the redistribution/incentive tradeoff.

Mirrlees' model is hypothetico-deductive. From a few assumptions, Mirrlees deduces certain properties of the optimal schedule and then the optimal schedule itself. To obtain the schedule, it is then sufficient to define the parameters. The parameters reflect the elasticity of labor supply (α), the concavity of the utility function (β), the weight of public spending to be financed by taxes (X/Z), and the distribution of individual productivity inequalities (σ).

The model is as follows:

[3] It would be easy to reject an income tax because of labor supply reaction and propose an optimal taxation of skills supposing "skills" are exogenous!



The transfer scale f() transforms taxable labor income (Ri) into disposable income (Rd), which is assumed to be equal to consumption (C) in the absence of savings. Taxable income Ri is equal to productivity, n, multiplied by hours worked (ht): Ri = n*ht. Productivity is assumed to be the only source of heterogeneity between individuals. Each individual with the same productivity n therefore makes exactly the same choice of hours worked and consumption. The ideal would be to tax productivity n, but this technology is not available (for an unspecified reason): the government is forced to use a tax on income rather than on productivity. Given this constraint, the utility of an individual with productivity n cannot be lower than that of an individual with lower productivity and income, since the individual with productivity n would then choose to reduce their working hours and income.

In theory, the optimal taxation paradigm allows for disentangling, in the event of disagreements on normative conclusions, what stems from empirical disagreements on σ and α from what stems from disagreements on normative parameters or the choice of social objectives.

Some of the initial elegance of the utilitarian model is lost when it comes to applying it. In the general model, what matters to society is defined solely by what matters to individuals according to how they count it. Thus, social utility is the unweighted sum of individual utilities, which are themselves the result of individual behavior: social preferences are thus formally nothing more than the sum of individual preferences. In a sense, there is no society. What we call "society" is nothing more than the unweighted sum of its components. In its application, this general principle tends to be lost. Individuals are assumed to have the same preferences. In the model and most of its successors, individuals therefore do not have their own preferences. "Social" preferences are thus chosen in practice by the modeler. This is not a problem in itself: making coherent proposals for public policy or social planning necessarily requires some form of structure and, therefore, something like social norms that cannot be deduced from observation. This is all the truer when it comes to redistribution and compulsory taxation, areas where preferences cannot be revealed by individual consumer choices.

The question of optimal redistribution is quite different from those dealing with modes of transport, for example. In transport economics, preferences for cars, buses, trams, or dedicated bike lanes can be revealed by observing behavior when the price and speed of each mode of transport vary. In theory, this makes it possible to deduce a willingness to pay for reduced travel time, independently for each mode of transport. In this case, gains and losses can be aggregated relatively easily because they can be observed, albeit indirectly, in euro equivalents. Planners can then choose whether or not to develop bus, tram, or bicycle transport based on willingness to pay, using a relatively straightforward cost-benefit approach,

even if they have to balance a certain level of efficiency on the one hand and the number and nature of winners and losers for each alternative on the other.

In the optimal taxation model, there is a choice between consumption and leisure, and there is therefore an *implicit price for leisure consumption* (at equilibrium, this price is equal to net wages after taxes), but unlike transport, public policy does not directly address these choices. Public policy attempts to raise taxes for redistribution and the financing of public goods; *individual choices do not directly inform these preferences*. In practice, individuals choose between C and L, and social choice is between the sum and distribution of C (consumption by the rich or consumption by the poor? private consumption or public services?). You cannot infer the social preference for equality from consumption choices.

The behavior of the taxpayer-worker, between leisure and consumption, does not inform the choice between consumption by the rich and the poor. This is therefore an important difference from a model where both individuals and society must choose between bus and train.

2.3. The initial conclusions

What does Mirrlees conclude in his 1971 article when he applies realistic parameter values to his model?

First, the article concludes that almost the entire population chooses to work (in fact, between 4 and 9% of the population chooses not to work according to the parameters used). In the model, if a person's marginal disutility of working is greater than the utility that additional consumption brings them, then that person chooses not to work. Although not many individuals choose not to work, many individuals choose to work only a few hours (both men and women, since individuals have no gender or family). A significant number of individuals work less than one-third of the time (which the author notes, but which is less emphasized as a result).

Second, the article concludes that, at the optimum, taxes remain low. There is not much redistribution.

Third, the article concludes that optimal effective marginal tax rates remain relatively low at the bottom of the distribution. The assumption of optimality of sharply declining marginal rates at the bottom of the distribution, and very high rates at the very bottom, would be rejected. At the top of the income scale, optimal *average* tax rates would be increasing, but optimal *marginal* rates would be slightly decreasing. The author does not rule out that a schedule with constant marginal rates, a linear schedule, is close to the optimal schedule.

Mirrlees' seminal article can therefore be seen as a refutation of the utilitarian justification for progressive tax rates in terms of marginal rates. There is a utilitarian argument for progressivity in terms of the decreasing marginal utility of consumption, but this does not consider the incentive argument. Taking both arguments into account, the optimality of a non-progressive

linear scale is not rejected. This does not prevent redistribution, but it could be achieved optimally through a lump-sum transfer and a flat tax (*basic income and flat tax)*.

Where do these conclusions come from? The explanation for the low rates at the bottom of the income scale is interesting. One might think that the optimal system should redistribute more heavily towards lower incomes due to a sharp decline in utility as consumption approaches zero. At the optimum, one should accept some efficiency losses at this level, and high effective marginal tax rates. Mirrlees explains that this is not the case and that effective marginal tax rates must remain low even at the bottom because some people do not work or reduce their working hours. Hence, the density of people at the bottom of the income scale is therefore higher. The optimal scale ten to keep marginal rates low *where people are concentrated*. This concentration effect offsets the redistribution effect; at the optimum, effective marginal tax rates are roughly constant according to declared income.

Mirrlees adds that some of these conclusions would be modified by a substantial change in the parameters. Optimal taxation is low because, in the initial assumptions, the tax burden needed to finance public spending is low. If this increases, the optimal tax increases, without fundamentally changing the *shape* of the tax scale[4].

Mirrlees shows that if inequalities in ability more than double, then at the optimum the number of people who do not work increases sharply (up to 20% in one of the detailed cases). Even in this case, marginal rates remain relatively low at the bottom of the scale to continue to encourage these people to work.

2.4. Extensions

Thousands of articles have used Mirrlees' framework, or an extension to the framework. Other objectives can be added (such as an insurance motive), constraints can be modified, it's possible to analyze the interaction of two instruments and two objectives. The framework can be formalized a bit differently, such as in Stiglitz (1982). The literature is so vast, that it is possible to show a paper for each objection one can raise to Mirrlees' initial article. This does not necessarily mean that the literature advances or is today more refined than in 1982. Since, the model is not empirically refutable, one could say that the literature advanced if it produced some form of consensus. This is not necessarily the case (see next section). But a consensus between those who work in the field would not be enough. Most worrying is that the field has not be able to provide any criteria for success. Emmanuel Saez and Gregory Mankiw do not agree on optimal taxation parameters (for example, the optimal shape of marginal tax rates). They do not agree either on the criteria that would distinguish right from wrong or just from unjust. This argument has a broader scope than Mirrlees' article or narrow framework, but we focus on that framework to make our point.

[4] On the contrary, for Heathcote and Tsujiyami (2021), the shape of the optimal tax and transfer schedule strongly depends on the level of revenue the government needs to raise: "*When we increase pressure on the government to raise revenue, the optimal marginal tax schedule becomes first flatter and then U-shaped, reconciling various findings in the literature*".

3. What have economists learned from this model and its extensions?

Has the Mirrlees model and its extensions taught us anything useful in the economic debate, i.e., not about the model itself, but about the desirable way to tax income in practice? Does the model achieve consensus among economists?

In a literature review ("*Optimal Taxation in Theory and Practice*"), Mankiw, Weintzierl, and Yagan (2009) [MWY hereafter] identify and comment on ten lessons from optimal taxation theory in recent decades. These lessons help answer the question we pose in this section.

Indeed, if these lessons are *new*, *non-trivial*, *substantial*, and economists *generally agree* with them, then we can say that the model has produced consensus.

3.1 The greater the inequality, the greater the redistribution must be

According to the first lesson of MWY, "*the marginal tax rate schedule depends on the distribution of capabilities*" and according to the fourth, "the optimal intensity of redistribution depends on income inequality." However, the lesson that "*the greater the inequality, the greater the redistribution*" is trivial within the model, since it follows relatively directly from the assumption of diminishing marginal utility. Utilitarians have defended progressive taxation based on this assumption since the end of the 19th century, so the result is not new. It follows that for the same concave utility function, the greater the initial differences, the more redistribution will increase social utility, because the sacrifice required of the rich to increase the income of the poor will then be relatively smaller.

The conclusion of the 1971 article goes further than simply stating that greater inequality leads to greater redistribution. Indeed, such a relationship can be found in a progressive socio-fiscal system, even with a constant tax scale. With a constant progressive tax scale, the greater the inequalities in the tax base, the greater the weight of socio-fiscal transfers. If inequality increases, more people will be subject to the top marginal income tax rate, and more people will receive the guaranteed minimum income, hence more redistribution. The 1971 article goes further than that. It concludes that the tax scale itself must be more progressive when inequality increases. MWY's formulation of the first lesson is consistent with this conclusion, but when it comes to testing this lesson in the empirical part of the article, the authors use *average* tax rates. We can therefore conclude that Mirrlees' argument has not become so intuitive fifty years later.

Tax scales may have become more progressive as inequality has increased, but this does not imply a normative agreement among economists. Also, in recent years, top marginal tax rates have fallen while income inequality has increased at the top of the scale (both trends have been documented by Piketty and co-authors, notably in *Capital in the Twenty-First Century)*.

In any case, Mirrlees' conclusion (the higher the inequality, the more progressive the tax scale should be) is difficult to apply in practice because *it is expressed in a counterfactual form,* which says nothing about the desirable level of redistribution. The counterfactual statement "*if* inequality increases, *then* the tax scale *should* be more progressive" is unlikely to resolve disputes over the progressivity of the tax scale in the absence of initial agreement. The statement is neither empirically testable nor applicable from a public policy perspective. It is merely a logical relationship in a model.

The same can be said for other normative conclusions expressed in different ways, for which there is consensus but *no substantial agreement*, such as: [the lower the average elasticity of labor supply, the higher the redistribution *should* be] or [the lower the social preference for equality, the higher the redistribution should be]. The latter statement is tautological, the former is trivial.

3.2 What is the optimal guaranteed minimum income?

Have studies on optimal taxation created consensus around the issue of guaranteed minimum income (the benefit paid to those with no primary income)? In 1971, and still today, the debate on public policies to combat poverty is quite different in the United States and the United Kingdom, or more generally in Europe. The question of assistance to the poor has preoccupied English economists since the beginning of political economy, with every economist commenting on the Poor Laws. The first Poor Laws date back to 1601, and the last ones were formally repealed in 1948 with the passage of the National Assistance Act, which aims to assist those who are not covered by the insurance system, as well as those who, although covered, have insufficient resources to meet their needs. The British law thus created universal coverage in terms of assistance based on need in 1948. As assistance is targeted at the poorest, it is not paid to people receiving income from work, which potentially raises an incentive issue. In the United States, the welfare system is not as advanced. Since 1964, able-bodied poor people have been eligible for the permanent food stamp program, but there is no income support covering the entire population, particularly people without children. The question of introducing such an income arose in the 1960s through the debate on negative income tax/universal income. Such a reform would not have the same effects in the United States, where it would introduce universal income support, as in the United Kingdom, where it would increase the financial incentives to return to work (see below).

In this heterogeneous institutional context, the optimal taxation paradigm allows discussing the social and tax system in abstract terms, without going into the details of national systems. What is the ideal tax scale in theory? On which national parameters should this scale depend in theory? Can different scales be objectively justified based on economic data on inequality and productivity distribution? These questions are legitimate and interesting and should help to objectify normative responses. But what has the optimal taxation paradigm taught us about this subject in particular?

In 1971, Mirrlees concluded that the higher primary inequalities are, the more individuals, at the optimum, choose not to work and must receive a high guaranteed minimum income. Consequently, if a technological shock increases inequalities in terms of individual productivity,

society must accept more people in voluntary unemployment and a higher guaranteed minimum income.

It is rare for economists to point out, based on an optimal taxation model, that it would be optimal for 5, 10, or 20% of the working population, male or female, to choose not to work if their disutility at work is lower than their productivity. In the event of a biased technological shock, the "race between education and technology" model is more often put forward: if technical progress is biased towards the most educated, rather than paying more money to more people on welfare (the model's response), the *mainstream* response today would be to educate more people, which is both egalitarian and efficient. “If productivity inequalities increase, then it is optimal for fewer people to work” is a conclusion of the Mirrlees model that is rarely mentioned.

The model does not seem to compel those who regularly use it to adopt recommendations with which they are uncomfortable. This is unfortunate: the advantage of a model whose main strength is its overall *consistency* would be to force or at least encourage its backers to adopt *all* its conclusions, not just those with which they are comfortable.

What has this model nudged economists to accept reluctantly? Just as a descriptive model can be judged by its degree of falsifiability (the more easily a model can be falsified, the more it tells us about the real world), a normative model can be judged by the conclusions it encourages to reject or adopt, despite prior inclinations.

The optimal level of income support or welfare is not generally debated as such, even though it is an output of the model. While there is public debate about the optimal form of marginal rates (see below), there is no real public debate using the optimal taxation paradigm about the optimal *level* of guaranteed minimum income. Yet the conclusion of the 1971 article focuses both on this relative level of guaranteed minimum income and on a complete schedule of marginal rates (which allows for a complete tax and transfer function). But in the 1971 article, and in many of its successors, the optimal schedule is presented as an optimal curve of *marginal rates,* a presentation that masks the optimal level of minimum income. An alternative presentation is to show income after transfer as a function of income before transfer. Marginal rates are then less visible, but distributional aspects are easier to identify. The presentation in terms of marginal rates emphasizes the incentive/behavioral aspect, the “economic” rather than distributional aspect.

Can the model nevertheless shed light on the optimal level of guaranteed minimum income? The commentary by Bourguignon (2001) on d'Autume's exercise leaves us skeptical about this possibility. It is interesting to reproduce long excerpts here:

> *Mirrlees' model is attractive in that it seems to touch on something fundamental in the issue of redistribution. It generated considerable enthusiasm during the 1970s. I think people believed at the time that the purely theoretical issues had been resolved and that, one way or another, econometrics would soon lead to practical applications. (…) Once we knew what the U(C, L) function representing agents' trade-offs between consumption and leisure actually was, all that remained would be to intelligently parameterize the social utility function in order to study the properties of the optimal tax schedule. The bad news of the last twenty or thirty years is that*

econometrics will probably not be able to estimate this function U(C, L) adequately, for both conceptual and practical reasons. (...) (p.66)

Firstly, there is the problem of the non-observability of the lowest levels of productivity. At the bottom of the distribution, there are people facing a marginal rate of around 100%—this is the case for those receiving minimum income support once the incentive period has ended. Clearly, such people should not be working. In fact, we observe that a large proportion of them do not work, while others are in paid employment. For the former, it is impossible to observe labor productivity. For the latter, the problem arises of the rationality of their behavior or, more precisely, the adequacy of the standard economic model of consumption-leisure arbitrage. (...) (pp. 68-69)

Aversion to inequality plays a significant role in explaining the shape of the optimal tax schedule. This is very clear in the article by A d'Autume. In our own work, a Rawlsian criterion leads to a guaranteed minimum income equal to 85% of average income, and of course to marginal tax rates that are decreasing but uniformly high, and consequently to a very low average monetary income. On the other hand, with the social utility function calibrated on the "revealed preferences" hypothesis, we obtain a much lower guaranteed minimum income (40% of average income when labor supply elasticity is low, by construction, and 13% when it is higher, 0.5 instead of 0.1), and marginal tax rates that decline very rapidly. (pp. 71)

Bourguignon thus highlights important conceptual problems with the optimal taxation model when it comes to thinking about public policy in practice. The conceptual and empirical problems lead to the conclusion that the optimal guaranteed minimum income varies between 13% and 85%, depending on the combination of normative and positive assumptions. This range of 1 to 6.5 is much wider than the range of policy proposals. To give an order of magnitude, in the 2022 French presidential election, Jean-Luc Mélenchon proposed a guaranteed minimum income of €1,060, and no candidate proposed lowering the amount of €575. The range of proposals among the most prominent candidates was thus in a ratio of 1 to 2. The model is therefore not very discriminating: it cannot be used to reject concrete proposals. From this point of view, it is unlikely to create a *new* consensus.

The conceptual problems highlighted by Bourguignon are serious. Focusing on those that relate more specifically to the level of the guaranteed minimum income, they concern: the practical and conceptual difficulty of estimating individual productivity for a population that does not always work; the corollary problem of a pure labor supply model when non-employment can be involuntary (unemployment and involuntary inactivity), the empirical observation of people who are employed even though they have no financial incentive to do so (and therefore do not conform to the theoretical model); the fact that guaranteed incomes are paid at the household level, whereas the labor supply concerns individuals (how can the potential productivity of the household be determined?). In concrete terms, the problems concern both the relevance of the concepts (theoretical simplifications) and the difficulty of empirically estimating the parameters of the model, the two being highly interrelated.

The theoretical advantage of the optimal taxation paradigm is that it distinguishes between facts and norms to conclude in terms of norms. The problem is that if the measurement of behavior is very imprecise, then differences in conclusions may arise from modeling errors in behavior or differences in norms. These modeling errors potentially affect both the parameters

of the empirical model and the form of the behavioral model: should it be a pure labor supply model? Individual or collective labor supply?

As Mirrlees noted as early as 1971, whether involuntary unemployment is included in the model has a potentially significant effect on the optimal level of guaranteed minimum income. In other words, the way in which empirical behavior is simplified in the model has a potentially *very* significant impact on normative conclusions.

The theoretical advantage of being able to distinguish between what comes from facts and what comes from norms in our disagreements thus disappears. In practice, due to a lack of adequate traceability, the model leads to confusion between errors on facts and norms, and ignorance about this confusion. The problem is exacerbated in this case because a pure labor supply model explains only a very small part of the differences in the labor market, particularly differences in unemployment (especially if we take the example of France since the 1980s).

As Mirrlees stated in 1971 : "*Since this conclusion is based on the analysis of an economy in which a man who chooses to work can work, I should not wish to see it applied in real economies*".

3.3 An optimal U-shaped or linear curve?

It is common to present a tax scale in marginal rates. This is how the government presents it to taxpayers. Taxation is said to be progressive for an individual (or group) if the (average) tax rate increases as income increases. Progressivity therefore depends on the distribution of marginal rates, but also on the amount of transfer to people outside the income bracket (via a guaranteed minimum income). Take the example of an economy where the average income is $2,000, the guaranteed minimum income is $1,000, and taxation is linear at a rate of 50%. The social and tax system can be considered highly redistributive despite the constant marginal tax scale (or linear scale).

Much of the literature on optimal taxation comments on this *effective* marginal tax rate curve. One of the conclusions of the 1971 article is that, at the optimum, *marginal* tax rates should *not be negative*: if an individual's taxable income increases by $100, their disposable income should not increase by more than $100. However, the *average* tax rate is negative at the bottom of the scale—individuals receive more in benefits than they pay in taxes.

In practice, the marginal rate curves observed in the 1970s (even more so than today) were often U-shaped because at the bottom of the income scale social benefits are degressive, while at the top, the income tax scale is progressive. It is this U-shaped curve that Mirrlees comments on.

Mirrlees' contribution is to show that this form may not be optimal and that the optimal curve *may* be close to linear taxation. The idea until then was that to be generous, high marginal rates were needed at the bottom of the scale, while the most productive should be encouraged with low rates. Hence, the optimal marginal rate slope should be decreasing. This is not the case in Mirrlees' model because it is necessary to continue to incentivize those at the bottom

of the scale. If there is one lesson from Mirrlees on which his heirs agree, it is that *the optimal tax schedule depends on where taxpayers are located on the taxable income scale*: it is optimal for marginal rates to be *relatively* lower (*ceteris paribus*) in places where more people are concentrated. It is one of the main lessons of this framework.

But once again, this lesson in a ceteris paribus variant is too general to draw public policy lessons. Solving the model should allow us to estimate the intensity of the contradictory effects. However, MWY (2009) conclude that the optimal schedule could be linear, while in a contemporary review of the literature, Brewer, Saez, and Shepard (BSS, 2010) conclude that it is U-shaped.

Where does the disagreement between MWY and BSS come from? Both articles comment on Mirrlees' original model. In this model, the difference in tax rates can only come from the distribution of abilities, the social utility function, or the elasticities of labor supply. According to MWY, it is not easy to understand where the difference (with Saez, 2001) comes from. This statement is disappointing: failing to create a consensus on the policies to be followed, one of the promises of optimal taxation is to shed light on the reasons for disagreement, but on reading MWY these are not so clear. MWY then attempt to reproduce Saez's results, which they succeed in doing by using a Pareto distribution of abilities. MWY show that using a different distribution law, a linear tax is close to the optimum. Following this reasoning, the disagreement therefore seems to be *empirical.* The fact that the disagreement concerns one of the empirical parameters is also disappointing. In an ideal scientific world, science should create consensus on facts (and behaviors), and policy disagreements should arise from disagreements on norms/values. Here, however, the disagreement seems to arise from the facts, or at least from something that resembles facts.

An important aspect of whether taxation should be linear or not is the treatment of very high incomes: should they be taxed at marginal rates that are higher, equal, or lower than those for lower incomes? According to lesson 2 of MWY, "*The optimal marginal tax rate may decrease for high incomes*." Strangely, this lesson contradicts lesson 3 ("*A proportional tax with a universal lump-sum transfer could be almost optimal*"), unless we take the conditionals seriously, in which case the real lesson is that we don't know. According to BSS, an estimate for the United Kingdom of the optimal marginal tax rate for high incomes is 56.6%, with a range of 50.4% to 64.5%. According to Saez (2001), the optimal top marginal rate is between 50 and 80%. These estimates depend mainly on the estimated elasticity of high incomes to taxation. This is because the weight of consumption allowed by these very high incomes in the social utility function is negligible (it is zero when incomes tend towards infinity). In fact, according to Saez, the optimal top marginal tax rate is the rate that maximizes revenue. In this case, utilitarian logic is abandoned, and the research question becomes entirely empirical/counterfactual: can tax revenues be increased by lowering (or raising) nominal tax rates? MWY's conclusion to this last question is that "*there is much debate about the true structure of income elasticities*." This seems like a disagreement over facts, but it is in the same time doubtful that with further knowledge about elasticities, the authors would converge on public policy recommendations.

At least BSS and MWY agree on one point: the result of an optimal marginal rate equal to zero (Stiglitz, 1982) has no practical relevance. Practical relevance being the only known type of relevance, it would be more accurate to conclude that the result is misleading because it never

applies. The initial result stems from the following intuition: if the government knows the capacity and preferences of the person with the highest income, and therefore their equilibrium income, then there is no cost to setting a marginal rate of zero above that income. This marginal rate then applies only to the additional effort that the person would make if the tax rate were to fall. The cost in terms of lost revenue is therefore zero, and the potential gain is positive if the individual responds to the incentive. For BSS, the gain is zero since the weight of consumption at this level is negligible. The theoretical result only applies to the wealthiest individuals and provided that their income is known in advance. In practice, however, the condition of perfect information is not met. Stiglitz relativized the relevance of this theoretical "result" from the outset, but this did not prevent the result from continuing to be discussed (within the discipline).

This "result" is symptomatic of a certain drift in the literature when it comes to the properties of the optimum, i.e., what can be said before any empirical resolution. Commenting on the properties of a normative model that has been chosen arbitrarily cannot provide us with direct information about the objectives that society should pursue. The optimal properties of an arbitrarily chosen model derive directly from the assumptions, in a tautological manner. The discussion is not without interest, but at this stage, it concerns the model itself and should mostly serve to refute the model. Indeed, used in this way, if the model does not have properties that are consistent with shared normative intuitions, this should argue against its use; on the contrary, if it behaves in accordance with shared intuitions, then it may be useful for discovering other intuitions.

The risk is to confuse the discussion of the mathematical properties of the model with the discussion of the desirable properties of the social and fiscal system. A striking example is that of *tagging* or optimal discrimination.

3.4 Tagging (discrimination) is always optimal

To what extent should taxation depend on personal characteristics other than income, such as gender, age, or size?

This debate is distinct from that on the ability to pay. In France and most other countries, progressive taxation is levied according to the ability-to-pay principle. To determine ability to pay, social and tax systems consider certain characteristics of households, such as the presence of children, dependent adults, or disability. For example, children are assumed to directly reduce the ability to pay.

The issue of "tagging" is different. The question that arises is: should taxation be based on age, gender, race, or height as proxies for productivity? This aspect was not directly addressed in the 1971 model, but the last paragraph of Mirrlees's conclusion points in that direction. Mirrlees points out that a better optimum could be achieved if it were possible to introduce a tax rate scale that depended solely on everyone's abilities and not on their efforts. From an efficiency standpoint, this approach has a clear advantage: if taxation does not depend on effort, it does not discourage people from working. Furthermore, in terms of equity, only

inequalities in endowment should be compensated, not those resulting from effort (according to the assumption made). The ideal would be a flat tax that depends only on individuals' productivity (to reduce inequality) and not on their actual income (so as not to discourage effort). Mirrlees suggests, for example, basing taxation on IQ[5] .

Following on from this, in an optimal taxation approach, economists suggest taxing differently according to age, gender, race, or size, which are unmanipulable variables that are correlated with income. Akerlof (1978) refers to this approach as "tagging." In this seminal article, he emphasizes that it is optimal for *single mothers* to receive a higher minimum income, even if this means facing higher marginal tax rates. The author generalizes and shows that it is optimal to do this for all less productive groups. This method reduces the cost of redistribution by reducing marginal rates on the most productive, thereby reducing the cost of disincentives. Akerlof then warns against this type of instrument, which poses three problems: (1) the incentive to be labeled, in this case by separating or falsely declaring that one lives alone; (2) the problem of horizontal equity; (3) the administrative cost.

Alesina et al. (2011) take up this line of argumentation and show that women's incomes should be taxed less than men's for two complementary reasons. Women have lower wages than men and are more sensitive than men to incentives in the labor market. Taxing women's income at lower rates than men's, therefore, has two positive effects: a redistributive effect and an incentive effect. Gender-based taxation would thus increase social welfare.

According to Mankiw and Weinziel (MW) (2010), taxation based on height would be optimal. The aim is to reduce inequality with as little impact as possible on production (in this case, the labor supply). Taxing income discourages work. It is better to tax size: since taller people have higher incomes on average, this reduces inequality, and since, unlike working hours, people do not choose their height (a strong assumption, since height must be measured), taxing height has no impact on production. We can therefore conclude that optimal taxation must take size into account. The authors conclude: "*if policies such as a height tax are rejected, then the standard utilitarian framework must fail to capture intuitive notions of distributive justice*," without it being clear whether this is in defense of taxation based on size or in criticism of utilitarianism[6].

Gilles Saint Paul is strongly opposed to gender-based taxation. Essentially, Saint Paul highlights the limitations of the arguments put forward by advocates of gender-based taxation (Alesina et al. 2007) and the Spanish policy proposal for gender-based taxation (i.e., lower taxes on women's income):

> *In one sense, the point that differentiating taxes by gender improves "welfare" is obvious. The reason is that non-differentiation is a special case of differentiation. One can always replicate*

[5] The problem with IQ is that it can be manipulated. Mirrlees minimizes this problem, however, suggesting that individuals will want to maximize their IQ scores for status reasons.

[6] The article is often cited as defending size-based taxation, yet the authors publicly deny this. The following passage also leaves room for doubt: *"Many readers will find the idea of a height tax absurd, whereas some will find it merely highly unconventional. We encourage all readers to consider why the idea of taxing height elicits such a response even though it follows ineluctably from a well-documented empirical regularity and the dominant modern approach to optimal income taxation. If the policy is viewed as absurd, defenders of this approach are bound to offer an explanation that leaves their framework intact*." (p.156)

> *the non-differentiating outcome by imposing the same taxes on both genders. If one relaxes that constraint, one has more instruments and "welfare" (i.e., whatever the "social planner" is supposed to maximize) must improve. All what Alesina et al. are saying is that if one does so, then the "optimal" tax on men is likely to be different from that of women, and that the former is likely to be higher, because male labor supply is typically more inelastic.*

If one has an objective to maximize, it is always better to have an additional instrument. In a short complementary text, Saint Paul (2008), adds:

> *All these are practical objections which suggest that the supposed gains from gender-biased taxation may turn out to be much smaller than the authors think. But my opinion is that such a proposal is both wrong and dangerous not because of these practical shortcomings but because it is about abolishing equality before the law. The authors do not seem to realize that they are contributing to the decay of our democratic institutions. (...)*
>
> *The question is: "How much of our civil rights are they willing to abolish to reach these Stalin-style planning objectives?"*

Interestingly, a philosophy with liberal foundations is criticized as totalitarian. To understand this, let us return to MW. They rightly argue that horizontal equity has no independent role in utilitarian theory. They then quote Kaplow (2002): "*Why would a society sacrifice potentially significant gains for its average member in order to preserve equal treatment among its members*?" But once we have opened this door, why stop at potentially significant gains? The optimum can only be achieved by discriminating on all known characteristics of the taxpayer, with a view to perfect discrimination, just as a program of perfect price discrimination allows a monopoly to extract all surplus. Kaplow's (2002) conclusion is that society should not take any considerations of fairness into account independently of utility:

> *In arguing that no evaluative importance should be given to notions of fairness, we are criticizing principles that give weight to factors that are independent of individuals' well-being or its overall distribution. (p. 465)*

The *welfarist* argument is hence taken to its logical conclusion: only individuals matter; *ultimately*, only their well-being matters; fairness should therefore have no *independent* weight in the drafting of laws. The question MW asks at the end of the article is the right one: can we really build a theory of progressive taxation based on utilitarianism? Before answering this question, we may ask whether the optimal taxation paradigm has influenced socio-fiscal systems.

4 Why the optimal taxation paradigm is unconvincing

The discussion so far has already shed some light on why the optimal taxation paradigm is less convincing than one might hope. This is not such a new conclusion. In fact, most economists who have worked with this paradigm have expressed serious reservations, either about the paradigm itself or about some of its standard results or their usual interpretation: Stiglitz

refutes the use of his non-taxation of capital optimality result[7]; Bourguignon relativizes the usefulness of the optimal redistribution paradigm (2001), Piketty and Saez (2013) question the underlying utilitarianism, as does MW, but for other reasons.

The question is: who today uses an optimal taxation model to defend a position that they did not already defend before using the model? Who has changed their mind because of a rigorous demonstration using the optimal taxation paradigm? If these sets are empty, then the model risks serving only as an argument from authority ("*optimal capital taxation is zero according to Stiglitz (1982)*[8]").

Hypotheses that are weaker than conclusions

A hypothetical-deductive model can only be convincing if individuals have stronger beliefs about the assumptions than about the conclusions. In this type of model, by design, the conclusions follow from the assumptions: the conclusions can only be convincing if there is prior agreement on the assumptions. The problem in this case is that individuals have stronger opinions about the model's conclusions (in terms of the tax rate) than about the assumptions (in terms of social preference for equality or the concavity of the utility function). Such a model is therefore unlikely to create consensus on the conclusions.

On the contrary, if the reader adopts a *reductio ad absurdum argument*, they may be led, in the event of disagreement on the conclusions, to reject the assumptions or the model itself.

One could use the model inversely by deducing social preferences from actual tax parameters (see LeGrand and Ragot, 2024). However, if you use these parameters to solve the optimal taxation model, you end up deducing an *ought* from an *is*. What society should do (tax optimally) is deduced from what it does, contrary to Hume's guillotine. Deducing preferences from what is done is conservative: if the tax system is supposed to reflect actual preferences, then why change it? Only efficiency arguments remain receivable, but then why use an equity/efficiency framework?

A model that must be judged on its relevance, but does not discuss it

A solvable normative model with a quantitative solution cannot, by definition, have universal validity. Its conclusions can only apply in a specific context, under restrictive assumptions. When considering public policy, it is important to understand the model's scope of validity. Mirrlees is very clear on this point in the conclusion of his article, writing explicitly that the real economy is very different from the one described in his assumptions and that he would therefore not like to see the conclusions of his model applied in the real world. The problem, he writes, is that there is involuntary unemployment due to labor supply exceeding demand. In this case, the guaranteed minimum income would have to be increased to a level at which

[7] "The Optimal Tax on Capital is Greater than Zero," INET Session in Honor of Tony Atkinson, New York, September 19, 2017. Stiglitz concludes: "It is wrong to conclude from Atkinson-Stiglitz 1976 that there should be no taxation of capital: there should be."

[8] Actually Stiglitz argued in 1982 that the optimal marginal taxation on the highest ability individual should not be zero but negative : "The widely discussed property of the optimal tax structure, that the most able individual faces a zero marginal tax rate, is only true if all individuals are perfect substitutes; in all other cases the highest ability individual should face a negative marginal tax rate."

the number of people choosing not to work is equal to the excess labor supply in the economy. These few lines in the conclusion are a rule of thumb. In any case, the presence of involuntary unemployment complicates the formal resolution of the model. The initial bet was that better models could be applied to the real world, but subsequent models seem to have the same weakness.

One problem is that the relevance and robustness of models to changes in parameters or modeling are rarely discussed. When they are, the discussion shows that the results are not very robust (see Bourguignon, 2001).

More generally, the question arises of the "academization" of normative economics as a branch of economics, and particularly its usefulness in public policymaking. Does normative modeling, which is inherently irrefutable, have a place in an empirical science? The risk is that it becomes metaphysics without reflexivity.

Intuitively, the transition from science to public policy on a subject such as taxation should lead to popularization, allowing for public deliberation and an open and pluralistic approach that considers a wide range of constraints and social objectives. However, the academization of normative economics, i.e., its publication in academic journals, tends to work in the opposite direction, sacrificing relevance to the consistency of a model with narrow objectives and limited behavioral responses. For example, the *height taxation* proposed by Mankiw and Weinzierl in *the American Economic Journal* is particularly absurd. Tax systems seek to rely on the ability to pay, but size is only very weakly correlated with the ability to pay. This taxation creates discrimination, and the assumption that individuals cannot influence their "tax size" is probably too strong. Finally, the problem identified (the disincentive effect of taxation on activity) is probably overestimated, while the problems surrounding the measurement of size are underestimated (and not seriously discussed).

The authors therefore propose a solution that is disconnected from the public policy perspective. The article is not interesting as a contribution to public policy but should be understood within the field of optimal taxation. The authors even admit that the article can be interpreted in two ways, either as a contribution to the field of optimal taxation or as a critique:

> *Some readers can take it as a small, quirky contribution aimed at clarifying the literature on optimal income taxation. Others can take it as a broader effort to challenge that entire literature.*

The exercise in style is such that the authors make a public policy proposal that can be interpreted seriously or as satire, depending on one's point of view... and the ambiguity is maintained by the authors themselves. The authors publish an article on a public policy proposal, but the reader does not know whether the authors support it or not. It is not that the authors list advantages and disadvantages and are unable to decide. On the contrary, they propose a model that allows a decision to be made (taxation of size combined with income is optimal), but the authors question whether the model is relevant. They invite us to reason by *reductio ad absurdum*: if we don't like the conclusion, then we don't like the model; if we like the conclusion, then we like the model. The article’s main conclusion is therefore about the model, not about the public policy.

A model that is not based on shared intuitions

When it comes to discussing taxation in practical terms for an audience that does not read academic journals, the optimal taxation paradigm is rarely used by authors publishing in this field. Emmanuel Saez is one of the most renowned authors in the field of optimal taxation. He has published numerous landmark articles on the optimal taxation of high incomes (Piketty, Saez, Stantcheva, 2014), on the optimal taxation of income with extensive labor supply responses (Saez, 2002), on the use of elasticities to derive optimal taxation (Saez, 2001), etc. However, in a recent book co-authored with Gabriel Zucman*, The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*, the optimal taxation paradigm is rarely, if ever, mentioned, even though the subject of the book would lend itself to it. Admittedly, the term "optimal" is used about fifteen times in the book, but in a sense that is quite different from the Mirrlees paradigm. The approach adopted in the book to determine the optimal top tax rate is summarized as follows:

> The optimal tax rate on the rich is simply the rate that raises the maximum possible revenue. It's not a controversial objective among economists. And it makes intuitive sense: everybody can agree that an extra dollar is much more valuable in the hands of a poor person than in Bill Gates's. Taxing the wealthy a bit more is not going to prevent them from affording good child care, but if raising rates allows those who serve them coffee or clean their houses to do so, too, it's worth it.

The argument might seem utilitarian. *"An extra dollar is much more valuable in the hands of a poor person than in Bill* Gates's" seems to follow from the idea of diminishing marginal utility. Nevertheless, "the rate that maximizes revenue" is purely an engineering problem, maximizing a real quantity (tax revenue). The argument is framed in such a way that it can be defended in terms of the right to quality childcare. What is written literally is that a society that guarantees child-care for all is preferable to a society that increases the consumption of billionaires. The authors continue: "*With this objective in mind, taxation becomes an applied engineering problem*." By "engineering problem", we mean here a problem that concerns only the behavior of a real quantity, the tax base. The aim is to maximize the taxes paid by the wealthiest. It is therefore not really a utilitarian program.

This may seem anecdotal, but it sheds light on the fact that when discussing tax systems for a broad audience, optimality in the sense of optimal taxation is rarely invoked today, probably because it is unconvincing.

From transparency to the argument from authority

The goal of transparency and clarification is rarely achieved by the optimal taxation paradigm, because solving the model is not easy, and because the conclusions are derived from both normative and behavioral parameters in a manner difficult to disentangle.

In practice, the paradigm of optimal taxation is too often used as an argument of authority: an academic article gives a "result" under very restrictive conditions, sometimes a theoretical result with no numerical application; then a public policy-oriented article recalls the main lines

of reasoning; and, finally, it is written in a press column that "economic analysis shows that". This is a general criticism that can be made of interventions aimed at the public, which tend to confuse "an economic analysis" with "the" economic analysis. The problem is compounded here by the fact that the analysis is not “economic”: it is mainly normative. It would therefore be more accurate to write "according to a normative analysis" or, more transparently, "according to my normative analysis," or more simply "in my opinion."

The drift in practice is linked to the structural problems already identified: the promise of breaking down disagreement comes up against a double conceptual and empirical imprecision, compounded by the veil of utility. The concepts are not always well defined, particularly the concept of "capabilities" (or "talents" or "productivity"), which is central to many articles. Added to the conceptual problem is the fact that empirical parameters, such as "*the elasticity of the tax base to marginal tax rates*" are known only with a great deal of imprecision. Finally, the paradigm maximizes utility, which is itself a concept and not a potentially observable quantity. As approximations interact and intersect in ways that are difficult to track, the promise of transparency of disagreement, decomposition between disagreements over norms and disagreements over facts, is difficult to keep.

Conclusion: opening the black box

If we define efficiency as the realm of the commensurable, then efficiency and fairness are incommensurable. This contradicts the paradigmatic assumption of optimal taxation, which seeks to quantify the trade-off between fairness and efficiency.

When individuals choose between two modes of transport with different speeds, prices, and levels of comfort, they make these characteristics commensurable (albeit not without problems). The case of taxation is very different because taxation is a purely social choice.

Ideally, the optimal taxation model could serve as a constraint and lead to the acceptance of counterintuitive conclusions (counterintuitive to the author). However, in practice, the model is not restrictive enough and often serves as a rhetorical argument or even, at times, as an appeal to academic authority.

To paraphrase Mirrlees' initial quote ("*a good way to govern is to agree on objectives, find out what is possible, and optimize*"), we can say that *a good way to govern is to discuss objectives, find out what is possible, deliberate on what is desirable, and decide*. In this framework, economists provide justifications for deliberating on what is possible, and, from a certain point of view, on what would be desirable, according to overall objectives. This calls for opening the black box and, probably, avoiding the utilitarian detour.

References

Akerlof G., 1978: « The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning », The American Economic Review, Vol. 68, No. 1, pp. 8-19.

Alesina A., A. Ichino, L. Karabarbounis, 2011: « Gender-Based Taxation and the Division of Family Chores », American Economic Journal: Economic Policy, Vol (3), n°2, pp. 1-40.

Arrow K., 1951: Social Choice and Individual Values, John Wiley & Sons.

D'Autume A., 2001 : "L'imposition optimale du revenu : une application au cas français », Revue française d'économie, 15-3, pp.3-63.

Blaug M., 1980: *The Methodology of Economics. Or how economists explain.* Cambridge University Press.

Bourguignon F., 2001 : « Commentaire sur l'article d'A. Autume », Revue Française d'Économie, 15-3, pp.65-73.

Brewer M., E. Saez et A. Shepard, 2010: "Means-testing and Tax Rates on Earnings", in *Dimensions of Tax Design*, The Mirrlees Review, Institute for Fiscal Studies, chaired by Sir James Mirrlees.

Buchanan J., 1964: "What should economists do?", *Southern Economic Journal,* Vol. 30, No. 3, pp. 213-222.

Edgeworth F., 1897: "The Pure Theory of Taxation". *The Economic Journal*, Vol. 7, No. 25., pp. 46-70.

Heathcote J. et H. Tsujiyama, 2021: " Optimal Income Taxation: Mirrlees Meets Ramsey", Optimal Income Taxation: Mirrlees Meets Ramsey. *Journal of Political Economy*.

Hill L., 1968: "A critique of positive economics", The American Journal of Economics and Sociology, Vol. 27, No. 3 (Jul., 1968), pp. 259-266.

Kaplow L., 2002: *Fairness versus Welfare*, Harvard University Press.

Le Grand, F., et Ragot, X., 2025: Optimal Fiscal Policy with Heterogeneous Agents and Capital: Should We Increase or Decrease Public Debt and Capital Taxes? *Journal of Political Economy*.

Mankiw, G. et M. Weinzierl, 2010: "The optimal taxation of height: A case study of utilitarian income redistribution", *American Economic Journal: Economic Policy* 2(1), 155–76.

Mankiw G., M. Weintzierl et D. Yagan, 2019: "Optimal Taxation in Theory and Practice", *Journal of Economic Perspectives*, Vol(23), n°4, pp. 147-74.

McCloskey D., 1983: "The Rhetoric of Economics", *Journal of Economic Literature*, Vol. 21, No. 2, pp. 481-5.

Mirrlees J., 1971: "An Exploration in the Theory of Optimum Income Taxation ", *The Review of Economic Studies*, Vol. 38, n°2, pp.175-208.

Mirrlees J., 1976: "Optimal tax theory: A synthesis", Journal of Public Economics, Vol (6), n°4, pp. 327-358.

Piketty T., E. Saez, 2013: “Optimal Labor Income Taxation” dans *Handbook of Public Economics, Vol. 5,* Auerbach, Chetty, Feldstein, Saez (ed.), Elsevier.

Piketty T., E. Saez, S. Stantcheva, 2014 : “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities”, American Economic Journal: Economic Policy, Vol(6), n°1, pp. 230-72.

Robbins L., 1932 : *An Essay on the nature and significance of economic science*. McMillan & co. Limited.

Saint Paul G., 2008: “ Against gender based taxation: Abolishing equality before the law is wrong and dangerous”, *Voxeu column*, 9 Feb 2008.

Seligman E., 1908 : *Progressive Taxation in Theory and Practice*, Princeton: American Economic Association;. London Swan Sonnenschein. 1908. pp. 334.

Sen A., 1979 : “Equality of what ?”, *Tanner Lectures on Human Values*. Stanford Univesity.

Simula L., A. Trannoy, 2007 : “ Imposition optimale sur le revenu et théorie des incitations : un chassé-croisé “, Regards croisés sur l’économie, n°1.

Slemrod J., 1990 : “Optimal Taxation and Optimal Tax Systems”, *Journal of Economic Perspective*, Vol. 4, n°1, pp. 157-178.

Stantcheva S., 2014 : “ Optimal Income Taxation with Adverse Selection in the Labour Market”, Review of Economic Studies, 81, pp. 1296-1329.

Stiglitz J., 1982 : “ Utilitarianism and horizontal equity : the case for random taxation”, *Journal of Public Economics*, Vol(18), n°1, pp. 1-33.

Stiglitz J. 1987, “Pareto efficient and optimal taxation and the new new welfare economics”, *Handbook of Public Economics*, Elsevier, pp. 991-1042.

ABOUT OFCE

The Paris-based Observatoire français des conjonctures économiques (OFCE), or French Economic Observatory is an independent and publicly-funded centre whose activities focus on economic research, forecasting and the evaluation of public policy.

Its 1981 founding charter established it as part of the French Fondation nationale des sciences politiques (Sciences Po) and gave it the mission is to "ensure that the fruits of scientific rigour and academic independence serve the public debate about the economy". The OFCE fulfils this mission by conducting theoretical and empirical studies, taking part in international scientific networks, and assuring a regular presence in the media through close cooperation with the French and European public authorities. The work of the OFCE covers most fields of economic analysis, from macroeconomics, growth, social welfare programmes, taxation and employment policy to sustainable development, competition, innovation and regulatory affairs.

ABOUT SCIENCES PO

Sciences Po is an institution of higher education and research in the humanities and social sciences. Its work in law, economics, history, political science and sociology is pursued through ten research units and several crosscutting programmes.
Its research community includes over two hundred twenty members and three hundred fifty PhD candidates. Recognized internationally, their work covers a wide range of topics including education, democracies, urban development, globalization and public health.
One of Sciences Po's key objectives is to make a significant contribution to methodological, epistemological and theoretical advances in the humanities and social sciences. Sciences Po's mission is also to share the results of its research with the international research community, students, and more broadly, society as a whole.

PARTNERSHIP