

Why we do not Tag to Tax: Concern for Horizontal Equity and its Implications

Vinzenz Zieseimer*

EUROPEAN UNIVERSITY INSTITUTE

Preliminary and Incomplete Draft

First Version: February 20, 2016

This Version: August 22, 2017

Abstract

The absence of certain tags is arguably the biggest difference between the practice and the theory of taxation today. Currently accepted Welfarist theories of optimal taxation suggest making taxes dependent on height, race, and gender. This paper establishes the precise condition under which a tag *must* be used by any Welfarist policy maker. In practice however, many tags are intuitively rejected on the basis of vague notions of Horizontal Equity, i.e. the equal treatment of equals. The paper formalizes this notion by minimally adjusting the classic Welfarist approach. In doing so, it achieves a near perfect correspondence between tags prescribed by theory and tags used in practice. This suggests that a range of academic proposals on differentiated taxation, such as gender-based taxation, simply do not conform to societal norms. Interestingly, the same applies to income-history dependent taxation, which has been the focus of the literature on dynamic optimal taxation.

1 Introduction

Arguably the biggest difference between the theory and practice of taxation is the use of tags (cf. Mankiw, Weinzierl, and Yagan (2009)). Theory, asking how a government can maximize

*_

the total welfare of its people, suggests that we should use any information available to optimize our tax system. As a result, tax theory suggests we should tax the tall more than the short, differentiate taxes by gender and ethnicity, and use genetic information on people's ability to earn money to determine how much taxes they should pay. It seems possible to construct examples that are even further from current practice: the government could randomize taxes, thereby generating information that can be used to optimize the tax system in later periods.

The main argument for tags like those based on height, despite the intuitive discomfort they invoke, has been the Pareto Principle. Indeed, it may be true that height-based taxation could make everyone better off in a Welfarist sense. As Mankiw and Weinzierl (2010), who discuss height-based taxation, put it: *"Nevertheless, if a nontrivial Pareto-improving height tax were possible, and if people both understood and were convinced of that possibility, it is our sense that most people would be comfortable with such a policy"* (p. 173). While this paper's preferred criterion takes a different view, it is important to investigate this argument further. Under which conditions is it precisely that tags facilitate a Pareto improvement? This paper provides novel insights. Tags are Pareto improving *if and only if* they identify a subgroup of the population for which there is a Laffer effect on average versus the optimal tax schedule without these tags. It immediately follows that if such a tag exists, then a Welfarist policy maker must use it. Thus, the scope for Pareto improvements by tagging is limited, but at the same time such tags are inherently linked to Welfarist policy making, or to any other method of policy assesment that respects the Pareto principle. If society is truly uncomfortable with certain types of discrimination, then perhaps a model in which the government strictly adheres to the Pareto principle cannot match reality.

While some of tags may intuitively feel ridiculous to some, it is still important to pin down why they are. Height-based taxation is often chosen as an example because of the strong intuitive response it provokes in many, but other tags feel much closer to home and are regularly discussed as policy options. Is there a fundamental difference between taxing based on height and taxing based on past income? If not, the entire literature on dynamic optimal taxation is at stake.

Why is it that we do not tag? The intuitive discomfort that comes with proposals such as height-based taxation is often related to concerns for Horizontal Equity, often defined as the principle that *equals should be treated equally*. This leaves much to be desired of course: what do we mean by equal? And by equal treatment? And what kind of tags does this principle then rule out? All this is the topic of this paper.

This paper considers the standard Mirrleesian environment, the most commonly used theoretical framework in the study of optimal taxation. It develops an adjusted Welfarist objective

criterion in which there is a minimal concern for Horizontal Equity. Notions of Horizontal Equity have been much criticized for their reliance on a ‘status quo’ or ‘natural state’ that is hard to envision in reality (see Kaplow (1989)). In addition, Horizontal Equity based objectives are often far from the more common Welfarist approach, and therefore lead to wildly different conclusions. This is undesirable, because other than the issue of tagging, the Welfarist objective does a good job at explaining why governments tax as they do. The adjusted Welfarist criterion that this paper develops relies neither on a ‘natural state’, nor does it diverge far from the original Welfarist criterion.

One previous author has taken up the question from which this paper borrows its title. Weinzierl (2014) introduces the idea that governments might in reality follow a *mixed objective*. While they care for standard Welfarist objectives, they are also concerned with horizontal equity, which Weinzierl (2014) formalizes as a principle of *Equal Sacrifice*. This approach differs in several ways from the one pursued here. First, this formalization requires a ‘natural state’, and therefore suffers from the most-heard critique of the principle of Horizontal Equity. Second, the resulting framework is far from the current theory of optimal taxation: rather than just reducing the role of some tags, it also alters the policy prescriptions on any other issue of taxation. The main advantage of the approach of Weinzierl (2014) is that it allows to uphold the Pareto principle, yet this is achieved by evaluating deviations from *Equal Sacrifice* by summed deviations in utility from an optimal *Equal Sacrifice* allocation, which is essentially a Welfarist evaluation of a Horizontal Equity objective. While such an approach is necessitated by the need to combine several objectives, it also means that *Equal Sacrifice* is more of an alternative bliss point than a principle of Horizontal Equity. As a result, the approach in Weinzierl (2014) prescribes a reduced use of tags such as height, but nevertheless still the use of it. And, as the theoretical results in this paper show, if these tags identify Laffer effects, then a criterion that respects the Pareto principle can never exclude them entirely. Administrative costs of introducing tags are the best candidate for bridging the remaining gap between theory and practice, yet it is questionable how large those costs would really be for easy-to-implement tags such as gender.

This paper takes more seriously the view of Horizontal Equity as a principle. Coincidentally, this also brings it closer to the way many laws are currently written: many forms of discrimination are forbidden in many countries, no matter the Pareto improvements they could bring. Of course, laws may simply be a rudimentary way of implementing a mixed objective. There are arguments for both viewpoints: while some research makes the case that mixed objectives underlie individual morality, Sausgruber and Tyran (2014) find that discriminatory taxes remain unpopular even when they deliver clearly dominant outcomes. Finally, the success of this paper’s approach may be its strongest argument: it turns out to have almost perfect predictive power.

The adjusted criterion proposed in this paper results in sharp predictions: it allows for some tags, but not for others, with little ambiguity. The result is a near perfect match with tags that are observed in reality, as well as with those that are not. This is particularly interesting in two respects: First, there has been a range of recent academic proposals on differentiated taxation, for example by gender. This paper suggests that these are unpopular simply because they do not conform to societal norms. Second, a recent literature on optimal taxation suggests making taxes dependent on the history of past incomes, but this can be ruled out by the same rationale this paper uses to rule out tags such as those based on height. Thus, if the theory presented in this paper is the best positive model of optimal taxation we currently have, then this has potentially grave implications for the normative dynamic optimal taxation literature: the objective function used there may be the wrong one, with great effect on the resulting policy recommendations.

Previous literature on tagging has focused on solving optimal taxation problems with tags under standard Welfarist criteria. I do not try to review this entire literature, but instead refer the interested reader to the following papers. Akerlof (1978) provides the first extensive treatment of the tagging problem. Immonen, Kanbur, Keen, and Tuomala (1998) explain how to solve a problem with tags in an otherwise standard Mirrleesian framework. Also related to this paper is any literature that derives welfare criteria from basic principles, seeks to understand how social preferences and political decisions arise from individual preferences, considers welfare criteria further removed from the classic Welfarist criterion, or considers taxation environments removed from the classic Mirrleesian equity-efficiency trade-off. Because this still includes entire fields of economics, I do not review these literatures here. In first instance, this paper should be seen as an undertaking in positive optimal taxation.¹

In what follows, section 2 introduces a Mirrleesian framework and discusses the relation between tags, Pareto optimality, and the Welfarist criterion. It then introduces an adjusted Welfarist criterion with concern for Horizontal Equity, and shows some implications for tagging. Next, it discusses how the adjusted criterion can be carried into a framework of dynamic taxation, and what implications this has. Finally, section 3 shows how the prescriptions of the adjusted criterion match the practice of taxation, and compares this to the prescriptions of a standard Welfarist framework. Section 4 concludes.

¹In this sense, the paper is related to the literature on ‘reverse optimal taxation’, which aims to find the social welfare weights implied by existing tax systems. See for example Hendren (2014) and the references therein.

2 Optimal Taxation, Tagging, and Horizontal Equity

I start by introducing a standard Mirrlees (1971) environment - which has one unobserved choice, one observed outcome, and one type of unobserved heterogeneity - with two added features: one extra dimension of observed heterogeneity - otherwise not connected, but perhaps correlated, with everything else - and observed heterogeneity in preferences. I then discuss the link between tags, Pareto optimality, and Welfarism. Next, I introduce an adjusted Welfarist criterion that respects Horizontal Equity, and discuss its implications for tagging. The environment and the paper's results intuitively generalize to many other static situations, so that I stick as close as possible to the most familiar setting. Generalizing to a dynamic setting is much less obvious, so that I finally devote considerable space to that undertaking.

2.1 Tags in a Mirrlees Environment

Denote individuals by index $i \in I$, which is a finite set. Each individual has an ability $\theta \in \Theta$ (a finite set of positive reals), which in combination with work time n results in production $y^i = \theta^i n^i$. Each individual has preferences over consumption $c \geq 0$ and work time $0 \leq n \leq 1$, which are ranked by a real-valued function $u^i(c^i, n^i)$ that is defined over the domain of the inputs. All u^i are assumed to be strictly increasing and concave in c , and strictly decreasing and convex in n .

As is standard in the Mirrlees setup, a social planner maximizes some criterion by setting tax policies. The planner observes neither θ^i nor n^i . Instead, he observes y^i , and some other individual characteristic, which is represented by a real number $\gamma^i \in \Gamma$. In addition, he has knowledge of the functions u^i and the distribution of abilities in the economy. Hence, the planner can set a function for taxes and transfers $T(y^i, \gamma^i)$. Consumption is finally determined by $c^i = y^i - T(y^i, \gamma^i)$.

From here on out, n_*^i denotes an optimal choice of agent i given the environment, and y_*^i and c_*^i the corresponding output and consumption. The problem of the social planner is completed by a welfare criterion which the planner maximizes subject to a constraint:

$$\sum_{i \in I} T(y_*^i, \gamma^i) = R. \quad (1)$$

Here, R represents some net revenue requirement of the government, which is a real number. For the remainder of this paper, we study the following generalized Welfarist objective function:

Definition 1 (Welfarist Criterion).

$$U(\{u^i(c_*^i, n_*^i)\}_{i \in I}) = \sum_{i \in I} w^i u^i(c_*^i, n_*^i),$$

$$\forall i : w^i > 0.$$

The utility functions must be interpreted cardinally and be known to the planner, in order to allow for a trade-off between the welfare of different individuals. Positive welfare weights must be attributed to all. That social preferences are a function of individual utilities only, without further objectives or significant restrictions, is the defining feature of the Welfarist approach.

Allowing for heterogeneity in individual utilities is conceptually difficult in the Utilitarian framework.² Formally, we simply assume the existence of cardinal u^i that are known to the planner. One interpretation of this framework that is equivalent for the purposes of this paper is as follows: we simply see utility functions as part of the social welfare function of the planner, so that their cardinality is set by the planner. At the same time, the planner knows the behavioral responses of the agents (at least at any relevant group level). This allows us to formalize the problem as if the u^i are known to the planner.³

How does a Welfarist social planner proceed? In this type of environment, the planner would ideally (first-best) set taxes based on individual ability. That way he could redistribute from those who are more able to those who are less so (depending of course on how much he cares about the agents). However, that information is not available to him. Thus, he must provide incentives for the more able. As Mirrlees (1971) shows, the optimal tax function then depends on the distribution of ability in the economy.

Tags provide additional information on the distribution of abilities, and therefore get the planner closer to his first-best. Because the tag splits the population into groups, which on average may have differing abilities, the planner can simply solve the optimal taxation problem per group, given some revenue requirement from the group. Next, the group revenue requirements are adjusted to minimize the social cost of meeting the total revenue requirement. Combining both steps yields a solution to the problem of the social planner where

²First, consider this example: two individuals provide the same n and receive the same c . Yet, they derive differential levels of welfare from this. Does a Utilitarian treat them differentially? The answer seems to be outside of the scope of common definitions of utilitarianism. (The same dilemma applies when welfare is differentially sensible to changes in n and c .) Second, suppose the cardinal u_i are not observed. Now how does the social planner elicit them? There does not seem to be an instrument for this.

³For the interested reader, the optimal Welfarist taxation implications of having heterogeneity in both u and ability heterogeneity are studied in Jacquet and Lehmann (2015).

groups may be treated differentially. The reader is referred to Immonen, Kanbur, Keen, and Tuomala (1998) for a more extensive exposition.

2.2 Pareto Implications

When does the planner make the tax schedule dependent on tags? When tags are entirely uninformative of ability, they are not used. As Weinzierl (2014) points out, the opposite does not need to hold for general Welfarist weights: welfare weights can be correlated with tags too, so that the net effect might be zero. However, two qualifying remarks should be made at this point. First, as is already pointed out in Mankiw and Weinzierl (2010), these would be knife-edge cases: one would have to construct welfare weights so that tags are excluded, which is hard to align with the individualistic principle underlying Welfarism. Second, and perhaps more importantly, I show below that some tags *must* be used by a Welfarist social planner. Thus, it is not true that the Welfarist criterion can always be *made to fit* basic notions of Horizontal Equity. But first, the following proposition establishes the link between tags and Pareto optimality.

Proposition 1 (Tags are Pareto improving if and only if they identify Laffer effects). *Suppose $T^*(y)$ maximizes the Welfarist Criterion if the planner would not observe a tag γ , and the planner observes γ . Then, there is a $T(y, \gamma) \neq T^*(y)$ that Pareto improves outcomes if and only if for some γ_k there exists a $T(y, \gamma_k)$ such that: $\forall y \in Y_{T(y, \gamma_k)} : T(y, \gamma_k) \leq T^*(y)$ and $\exists y \in (Y_{T^*(y)}^{\gamma_k} \cup Y_{T(y, \gamma_k)/T^*(\gamma)}^{\gamma_k}) \exists k : T(y, \gamma_k) < T^*(y)$, while $\sum_{\{i \in I : \gamma^i = \gamma_k\}} (T(y_*^i, \gamma^i) - T^*(y_*^i)) \geq 0$.*

Here, for generic tax schedules T and T' , $Y_T = \{y : \exists i (y_T^i = y)\}$, $Y_T^{\gamma_k} = \{y : \exists i (\gamma^i = \gamma_k) \cap (y_T^i = y)\}$, and $Y_{T/T'}^{\gamma_k} = \{y : \exists i (\gamma^i = \gamma_k) \cap (y_T^i = y) \cap (y_{T'}^i \neq y)\}$, where y_T^i denotes the agent's optimal choice is with respect to tax function T .

Proof. I start with the *only if* part. First, observe that for a Pareto improvement we require at least one agent to be better off, and no one worse off:

$$\begin{aligned} \exists i : u^i(\theta^i n_{T(y, \gamma_k)}^i - T(\theta^i n_{T(y, \gamma_k)}^i, \gamma^i), n_{T(y, \gamma_k)}^i) &> u^i(\theta^i n_{T^*(y)}^i - T(\theta^i n_{T^*(y)}^i, \gamma^i), n_{T^*(y)}^i), \\ \forall i : u^i(\theta^i n_{T(y, \gamma_k)}^i - T(\theta^i n_{T(y, \gamma_k)}^i, \gamma^i), n_{T(y, \gamma_k)}^i) &\geq u^i(\theta^i n_{T^*(y)}^i - T(\theta^i n_{T^*(y)}^i, \gamma^i), n_{T^*(y)}^i). \end{aligned}$$

We cannot raise any effective tax rate. This tax rate would be paid by two types of agent: those for whom the optimum is the same under both schedules, and those for whom it is different. For those for whom it is the same, the effective tax rate has risen (and welfare is strictly increasing in consumption). Those for whom it is different would have also preferred the old tax schedule, and the old tax schedule was available to them under the original tax system in which it was as good as their optimum at best (if they were indifferent). Thus, raising effective tax rates ($Y_{T(y, \gamma_k)}$) makes some agent worse off.

To make some agent with generic characteristic γ_k better off, we can either lower his tax $T(y, \gamma_k)$ at his optimal choice ($Y_{T^*(y)}^{\gamma_k}$), or lower his tax at another choice which then becomes his optimal choice while the previous is no longer optimal ($Y_{T(y, \gamma_k)/T^*(\gamma)}^{\gamma_k}$).

Finally, lowering some effective taxes while not increasing any is only possible if this increases tax revenues or keeps them the same: if we would not have $\sum_{\{i \in I: \gamma^i = \gamma_k\}} (T(y_{T(y, \gamma_k)}^i, \gamma^i) - T^*(y_{T^*(y)}^i)) \geq 0$, then the budget constraint would be violated. Lowering some effective taxes (which is the same as lowering the tax rate at the relevant income level) and increasing or maintaining revenues in the process is commonly referred to as a Laffer effect.

The *if* part is straightforward. If we implement the $T(y, \gamma_k)$ for which the condition holds, then none are worse off and some agent is better off, because the same choice would result in higher consumption and welfare is strictly increasing in c . \square

Proposition 1 shows how one can improve upon an already optimal system by introducing tags: essentially, one needs to identify a group for which, on average, a lower tax rate than the previously optimal one results in higher taxes paid. This is similar to how one identifies Pareto improvements in any given suboptimal tax schedule without the use of tags: one identifies Laffer effects, raising the same or more taxes using lower effective tax rates.⁴ Whether Laffer effects are present very much depends on the tag to be used. Do we believe there is some height group for which a Laffer effect is possible on average? Mankiw and Weinzierl (2010) do report a Pareto improving use of height as a tag in their calibrated procedure, but it is very small. Absent empirical evidence, we can only speculate.

Suppose that the finding in Mankiw and Weinzierl (2010) is correct, and height-based taxation can indeed be used to achieve a Pareto improvement. Then, recasting the proposition above in the manner of Kaplow and Shavell (2001): *Any otherwise Welfarist method of tax policy assessment that does not discriminate on the basis of height violates the Pareto Principle*. In other words, the relation between tagging and the Welfarist criterion is inherent to Welfarism. This notion is formalized in the following corollary.

Corollary 1 (Welfarism implies using any tags that identify Laffer effects). *Suppose the planner maximizes the Welfarist Criterion and observes γ , and suppose that for some γ_k there exists a $T(y, \gamma_k)$ such that: $\forall y \in Y_{T(y, \gamma_k)} : T(y, \gamma_k) \leq T^*(y)$ and $\exists y \in (Y_{T^*(y)}^{\gamma_k} \cup Y_{T(y, \gamma_k)/T^*(\gamma)}^{\gamma_k}) \exists k : T(y, \gamma_k) < T^*(y)$, while $\sum_{\{i \in I: \gamma^i = \gamma_k\}} (T(y_{T(y, \gamma_k)}^i, \gamma^i) - T^*(y_{T^*(y)}^i)) \geq 0$. Then for the resulting optimal*

⁴Werning (2007) discusses conditions under which a tax system is Pareto efficient, but only for separable and homogeneous preferences. He then generalizes his conditions to the case of tagged groups. In essence, he finds conditions under which Laffer effects can be found. Here I show, in a general environment, what type of Laffer effects are needed so that Pareto improvements can be achieved in an optimal taxation system when one introduces tags.

tax schedule $T^*(y, \gamma)$ we have $T^*(y, \gamma) \neq T'(y)$ for any function T' that does not depend on γ .

Proof. Suppose $T^*(y, \gamma) = T'(y)$ for some T' that does not depend on γ . Obviously, we must have $T'(y) = T^*(y)$, i.e. it is a tax function that solves the planner's problem without knowledge of the tag. Now, according to Proposition 1 a Pareto improvement is possible, given what we suppose on $T(y, \gamma_k)$. But then, the Welfarist Criterion cannot have been maximized because each agent has positive weight, and we have

$$\begin{aligned} \exists i : u^i(\theta^i n_{T(y, \gamma_k)}^i - T(\theta^i n_{T(y, \gamma_k)}^i, \gamma^i), n_{T(y, \gamma_k)}^i) &> u^i(\theta^i n_{T^*(y)}^i - T(\theta^i n_{T^*(y)}^i, \gamma^i), n_{T^*(y)}^i), \\ \forall i : u^i(\theta^i n_{T(y, \gamma_k)}^i - T(\theta^i n_{T(y, \gamma_k)}^i, \gamma^i), n_{T(y, \gamma_k)}^i) &\geq u^i(\theta^i n_{T^*(y)}^i - T(\theta^i n_{T^*(y)}^i, \gamma^i), n_{T^*(y)}^i). \end{aligned}$$

□

Because the corollary essentially just combines the proposition above with the observation that Welfarism respects the Pareto criterion, it extends to any other objective that respects the Pareto criterion. This includes the Rawlsian leximin criterion, or the mixed objective in Weinzierl (2014).

The main argument for Welfarism, and the main objection against horizontal equity concerns in optimal taxation, has always been the potential to violate the Pareto principle: if we can make everyone better off, why would we not do so? The above results raises the reverse question: if we really object to certain forms of discrimination, should we adhere so strictly to Welfarism and the Pareto principle? While Pareto optimality has always been a benchmark principle for economists, the same cannot be said for philosophers of economics.

An alternative to the principle of Horizontal Equity is to consider mixed objectives, where there is both concern for horizontal equity and Pareto optimality. It may well be that we should adjust our criterion to allow exception to horizontal equity when Pareto improvements are possible. This is hard to determine as a result of positive optimal taxation, since an example of a diffuse tag that is clearly Pareto improving (and of significant size) seems to be missing. A similar point can be made with regards to current legislation on discrimination. Are laws that forbid certain types of discrimination based on fundamental principles, or do they just rigidly implement one part of a mixed objective when it clearly dominates? In what follows, I take the former view.

2.3 Horizontal Equity

I now make an attempt to formalize the intuitive aversion to tags such as those based on height. In doing so, I stay as close as possible to the Welfarist objective, while at the same time introducing a minimal notion of horizontal equity.

The concept of horizontal equity, which is often casually introduced as *the equal treatment of equals*, has gained little acceptance so far. Despite its intuitive appeal, the concept suffers from the major critique that its operationalization seems to require a choice of some natural state against which to compare allocations, a point forcefully made by Kaplow (1989). Indeed, the previous literature has made such choices. For example, to introduce the notion of Equal Sacrifice, an objective that competes with the Utilitarian criterion, Weinzierl (2014) needs some starting point from which to calculate an individual's sacrifice. He chooses the allocation in which there is no government intervention, but at the same time notes: “A well-known conceptual issue with the idea of the *laissez-faire* allocation is that any economy is, in reality, inseparable from the government and state institutions that taxes fund. The *laissez-faire* allocation is, therefore, not well-defined, because [the absence of taxation] implies a very different economy than the status quo” (p. 137). The seminal paper of Fleurbaey and Maniquet (2006) similarly defines a *laissez-faire allocation*, which is equivalent to one without taxation, and uses it to introduce a notion of horizontal equity. In the adjusted Welfarist criterion below, I avoid this issue entirely by defining horizontal equity with respect to other individuals only, irrespective of the allocation.

For given tax function T , denote by $x_{i'}^i$ the level of choice variable x for agent i when agent i chooses the level that is optimal for agent i' (so that $x_i^i = x_*^i$).

Definition 2 (Welfarist Criterion with Horizontal Equity).

$$\begin{aligned}
U(\{u^i(c_*^i, n_*^i)\}_{i \in I}) &= \sum_{i \in I} w^i u^i(c_*^i, n_*^i), \\
\forall i : w^i &> 0, \\
\forall i, i' : (\theta^i n_{i'}^i = y_{i'}^i) &\implies u^i(y_{i'}^i - T(y_{i'}^i, \gamma^i), n_{i'}^i) \geq u^{i'}(y_*^{i'} - T(y_*^{i'}, \gamma^{i'}), n_*^{i'}).
\end{aligned}$$

The Welfarist criterion has been further restricted: Suppose a generic agent i could produce the same as agent i' using the same input. Then in doing so he is to obtain at least the level of welfare that agent i' obtains.⁵

The restriction is purposefully minimal. An agent of higher ability does not have the right to the same T as one of lower ability: the criterion only applies to agents who have the same ability.⁶ Agents of the same ability are not necessarily treated the same: someone who cannot provide the same level of n as others with his ability is not subject to this criterion: it is only defined for those who can imitate both behaviour and production. At last, individuals

⁵The extra restriction also reduces the relevant space of weights w , but there is no need to explore this further.

⁶One could turn the extra condition in the adjusted criterion into a restriction on rank reversals, by replacing $(\theta^i n_{i'}^i = y_{i'}^i)$ by $(\theta^i n_{i'}^i \geq y_{i'}^i)$. Much of the following would remain unchanged.

with clearly distinct preferences (whether cardinal or relative) may be treated differentially. Nevertheless, this minimal criterium prescribes *equal treatment of equals*, as the following proposition shows.

Proposition 2 (Equal Treatment of Equals). *Suppose, for two agents i and i' , we have $u^i(\cdot) = u^{i'}(\cdot) = u(\cdot)$ and $\theta^i = \theta^{i'}$. Then we must have $T(y_*^i, \gamma^i) = T(y_*^{i'}, \gamma^{i'}) = T(y_*^i)$ and $y_*^i = y_*^{i'}$, i.e. taxes paid at the optimal choice of either agent do not depend on individual characteristics γ^i and $\gamma^{i'}$.*

Proof. Since we have that $(\theta^i n_{i'}^i = y_{i'}^i)$ and $(\theta^{i'} n_i^{i'} = y_i^{i'})$, we must have both $u(y_{i'} - T(y_{i'}, \gamma^i), n_{i'}) \geq u(y_{i'} - T(y_{i'}, \gamma^{i'}), n_{i'})$ and $u(y_i - T(y_i, \gamma^{i'}), n_i) \geq u(y_i - T(y_i, \gamma^i), n_i)$. (The equivalent notation $x_i^i = x_*^i$ is used. Superscripts are irrelevant other than for γ , as we are dealing with two otherwise identical agents.) At the same time, because n^i is an optimal choice for agent i and $n^{i'}$ is an optimal choice for agent i' given T , we also have $u(y_{i'} - T(y_{i'}, \gamma^{i'}), n_{i'}) \geq u(y_i - T(y_i, \gamma^{i'}), n_i)$ and $u(y_{i'} - T(y_{i'}, \gamma^i), n_{i'}) \leq u(y_i - T(y_i, \gamma^i), n_i)$. Combining results in equality of all four evaluations of u . Then, $T(y_i, \gamma^i) = T(y_i, \gamma^{i'}) = T(y_i)$, $T(y_{i'}, \gamma^i) = T(y_{i'}, \gamma^{i'}) = T(y_{i'})$, and thus $y_i = y_{i'}$. \square

A direct consequence of this is that certain tags cannot be used: no matter how informative a tag is of a group's abilities, if there are *equals* everywhere along the optimal choices across groups which are identified by a tag, then this tag cannot be used. By the example of height: if for every input level chosen there is at least one tall and one otherwise equal short person choosing that level, then height is ruled out as a tag.

Definition 3 (Diffuse Tags). *Call Diffuse Tags any characteristics represented by γ , such that whenever there exists an agent, say i , who holds some level of γ , say γ^i , then there exists another agent, say i' , such that $\gamma_i \neq \gamma_{i'}$, $u^i(\cdot) = u^{i'}(\cdot)$, and $\theta^i = \theta^{i'}$.*

Proposition 3 (Exclusion of Diffuse Tags). *A planner who chooses amongst allocations according to the Welfarist Criterion with Horizontal Equity will not use Diffuse Tags. I.e. we have $T(y, \gamma) = T(y)$ for any observed level of y when γ is diffuse.*

Proof. This follows directly from Definition 3 and Proposition 2. Take a generic level of y that is optimal to some agent i with characteristics represented by γ^i : y_*^i . Then according to Definition 3, there exists another agent i' with optimal level $y_*^{i'}$, $\gamma_i \neq \gamma_{i'}$, while $u^i(\cdot) = u^{i'}(\cdot)$ and $\theta^i = \theta^{i'}$. According to Proposition 2, we must have $T(y_*^i, \gamma^i) = T(y_*^{i'}, \gamma^{i'}) = T(y_*^i)$. Since we had picked a generic level of y that was optimal to some agent, this holds for all observed y . \square

Of course, one can argue that in practice no tag is entirely diffuse: when the cardinality of the set of observed abilities is large enough compared to that of the set of individuals, then one will not find precise equals at all. However, the government in practice is not perfectly knowledgeable of the distribution of ability in the economy. It sets taxes in advance, after which there might in theory be equals anywhere. Another way to put this for the example of height is as follows: the support of the distribution of abilities is *ex ante* the same for all height groups. Thus, in practice the state cannot use height as a tag because at any level of production there *could* be several people of any height: height is a diffuse tag.

Note that this result does not depend on our interpretation of the u^i . Even if preferences are correlated with diffuse tags, they are ruled out. At the same time, if there are surely no equals somewhere in the sense of the Horizontal Equity restriction, then a tag can be used at least there.

2.4 Dynamic Optimal Taxation

So far I have discussed a static Mirrleesian environment. A growing literature discusses optimal taxation over a life-cycle, say an agent lives from age 0 to age T , where the nature of the problem is different because an agent's earnings ability $\theta^{i,t}$ may change over time, and the agent can self-insure through asset holdings a :

$$a^{t+1} = a^t + y^t - T^t(y^t, a^t, \gamma^t, \{y^s\}_{s=0}^{t-1}, \{a^s\}_{s=0}^{t-1}) - c^t,$$

where a^0 is given. Asset holdings are typically considered observable to the planner. The planner maximizes a standard dynamic Welfarist criterion, as presented below, by taking into account the history of incomes of an individual. This is because past incomes contain information on current and future ability, unless the ability process is independent over time.

Definition 4 (Dynamic Welfarist Criterion).

$$U(\{\sum_{t=0}^T u^{i,t}(c_*^{i,t}, n_*^{i,t})\}_{i \in I}) = \sum_{i \in I} w^{i,t} (\sum_{t=0}^T u^{i,t}(c_*^{i,t}, n_*^{i,t})),$$

$$\forall t \forall i : w^{i,t} > 0.$$

Introducing horizontal equity into this criterion can be done in several ways, which is discussed below. I propose the following. (For given tax function T , denote by $x_{i',t}^{i,t}$ the level of choice variable x for agent i of age t when agent i chooses the level that is optimal for agent i' of age t (so that $x_{i,t}^{i,t} = x_*^{i,t}$).)

Definition 5 (Dynamic Welfarist Criterion with Horizontal Equity.).

$$\begin{aligned}
U(\{\sum_{t=0}^T u^{i,t}(c_*^{i,t}, n_*^{i,t})\}_{i \in I}) &= \sum_{i \in I} w^{i,t} \left(\sum_{t=0}^T u^{i,t}(c_*^{i,t}, n_*^{i,t}) \right), \\
\forall t \forall i : w^{i,t} &> 0, \\
\forall t \forall i, i' : (a^{i,t} = a^{i',t}) \& (\theta^{i,t} n_{i',t}^{i,t} = y_{i',t}^{i,t}) &\implies \\
u^{i,t}(y_{i',t}^{i,t} - T^t(y_{i',t}^{i,t}, a^{i,t}, \gamma^{i,t}, \{y^{i,s}\}_{s=0}^{t-1}, \{a^{i,s}\}_{s=0}^{t-1}) - a_{i',t}^{i,t+1} + a^{i,t}, n_{i',t}^{i,t}) &\geq \\
u^{i',t}(y_*^{i',t} - T^t(y_*^{i',t}, a^{i',t}, \gamma^{i',t}, \{y^{i',s}\}_{s=0}^{t-1}, \{a^{i',s}\}_{s=0}^{t-1}) - a_*^{i',t+1} + a^{i',t}, n_*^{i',t}). &
\end{aligned}$$

This seems to be the most straightforward extension of the idea of horizontal equity to a dynamic environment. An agent who at some age starts off with the same wealth, behaves the same (both in his labor and asset-consumption choice), and produces the same as another agent of that age, should not be worse off than that other agent. The following corollary shows an implication of this choice.

Corollary 2 (Taxes do not depend on diffuse tags, including Past Incomes or Assets).

Assume that whenever there exists an agent i of age t , with history and characteristics $\{\gamma^{i,t}, \{y^{i,s}\}_{s=0}^{t-1}, \{a^{i,s}\}_{s=0}^{t-1}\}$, then there exists another agent of age t , say i' , such that $\gamma^{i,t} \neq \gamma^{i',t}$ but $a^{i,t} = a^{i',t}$, $u^{i,t} = u^{i',t}$, and $\theta^{i,t} = \theta^{i',t}$. What holds true for γ for some two agents, also holds for $y^{i,s}$ when $0 \leq s \leq t-1$, and $a^{i,s}$ when $0 \leq s \leq t-1$.

Then we have $T^t(y^t, a^t, \gamma^t, \{y^s\}_{s=0}^{t-1}, \{a^s\}_{s=0}^{t-1}) = T^t(y^t, a^t)$ for any observed level of y^t and a^t .

Proof. This follows from straightforward re-application of the proofs of propositions 2 and 3. \square

The same critique as in the static case applies. Will we really find two such agents? Again I argue that if the state sets taxes in advance, not knowing the distribution of ability quite perfectly, then the result applies. Would we ever expect to find two otherwise equal agents with the same asset levels at the beginning of a period, but different income and asset histories? At the very least this is possible because of changing preferences for savings. In addition, note that many relevant elements of the real world have not been incorporated in this framework. These include risk of non-diversified financial investments, unexpected expenditures on health (whether own or those of relatives), bequests, and transfers to children.

Under the specification chosen here, age dependent taxation is still permitted. Age-dependent taxation has often been presented as a simplifying short-cut to get most of the benefit of dynamic taxation without the complicated policies that come with it. Examples include Weinzierl (2011), Farhi and Werning (2013), and Stantcheva (2015). Under the criterion we present here, such proposals actually become optimal policies.

Other specifications that carry the principle of Horizontal Equity from the static setting into a dynamic one are possible. For example, we could compare all ages at every period, instead of only those of the same age. In that case, age-dependent taxation would have been excluded. We could also compare over entire life-cycles, in which case both age-dependent and income-history dependent taxation would be acceptable. Comparing life-cycles would take some notion of life-cycle effort and production in order to define horizontal equity if we do not just want to consider different income paths incomparable (in which case Horizontal Equity would have no bearing on past incomes and assets). It is not clear how to set taxes by period in that case, and the result would still be far from the prescriptions of the dynamic optimal taxation literature. The version proposed here seems closest in nature to the original criterion, and coincidentally produces the best explanation of the policies we observe. While it is somewhat harder to operationalize Horizontal Equity in a dynamic environment, one conclusion seems worth drawing: If we accept horizontal equity as relevant for taxation, then this has large implications for the dynamic optimal taxation literature.

2.5 Solving Optimal Taxation Problems under Horizontal Equity

How would one solve optimal taxation problems under the criteria suggested here? This clearly depends on the tags one wants to introduce to the environment. Nevertheless, the case for diffuse tags is clear: they should not be considered by the planner. There are generally two ways of achieving this. The first is to not make them part of the planner's information set. In that sense, many papers that have been written on the topic of optimal taxation simply remain valid in the presence of diffuse tags, under the adjusted criterion. The second is to take the tag as being randomly distributed over the population, so that it contains no information value.

This is particularly relevant for dynamic optimal taxation, where Albanesi and Sleet (2006) provide some results for the case where abilities evolve independently over time. That problem is equivalent to solving the taxation problem with abilities that are dependent over time (but diffuse), but under a criterion with horizontal equity that disregards age. The first alternative specification suggested above implies precisely that.⁷ Unsurprisingly, Albanesi and Sleet (2006) find that optimal taxation can be implemented without making taxes dependent on past income or past assets.

⁷The same cannot be said about numerical results, where the evolution of ability should still resemble the actual process.

3 Empirical Implications

In this section, I show that introducing horizontal equity as defined in this paper results in sharp implications that match reality near perfectly. Notably, the match between the criterion’s implications and reality is not approximate, but precise, in two ways: First, there is a close link between tags that are used and not used under the criterion and in reality. Second, excluded tags are not just used to a lesser extent, as for example in Weinzierl (2014), but entirely excluded on the basis of principle.

Static				
<i>Tag</i>	<i>Observed</i>	<i>Welfarist</i>	<i>HE</i>	<i>Reference</i>
Income	Yes	Yes	Yes	Mirrlees (1971)
Consumption	Yes	Yes	Yes	Atkinson and Stiglitz (1976)
Unemployment	Yes	Yes	Yes	Saez (2002)
Disability	Yes	Yes	Yes	Diamond and Mirrlees (1978)
Blindness	Yes	Yes	Yes	Weinzierl (2014)
Household Form	Yes	Yes	Yes	Kleven, Kreiner, and Saez (2009)
Children	Yes	Yes	Yes	Domeij and Klein (2013)
Mortgage Interest	Yes	Yes	Yes	–
Charitable Contributions	Yes	Yes	Yes	Blumkin and Sadka (2007)
Health Expenditure	Yes	Yes	Yes	–
Gender	No	Yes	No	Alesina, Ichino, and Karabarbounis (2011)
Height	No	Yes	No	Mankiw and Weinzierl (2010)
Race	No	Yes	No	Blumkin, Margalioth, and Sadka (2009)
Genetics	No	Yes	No	Logue and Selmrod (2008)
Other Diffuse Tags	No	Yes	No	Mankiw, Weinzierl, and Yagan (2009)
Dynamic				
<i>Tag</i>	<i>Observed</i>	<i>Welfarist</i>	<i>HE</i>	<i>Reference</i>
Income	Yes	Yes	Yes	Golosov, Kocherlakota, and Tsyvinski (2003)
Past Income	No	Yes	No	Kocherlakota (2005)
Assets	Yes	Yes	Yes	Golosov, Kocherlakota, and Tsyvinski (2003)
Past Assets	No	Yes	No	–
Income Averaging	No	Yes	No	Vickrey (1939)
Age	Yes	Yes	Yes	Weinzierl (2011)
Education	No	Yes	No	Bohacek and Kapicka (2008)
Income-contingent Loans	Yes	Yes	Yes	Stantcheva (2015)

Table 1: Tagging under Horizontal Equity

Table 1 lists a large number of tags that are either part of the US or other countries’ income

tax code, or have been discussed in the academic literature, or have been the subject of public debate. Dynamic tags have been separated from those that are also relevant in a static setting. For each tag, I indicate whether it is used in the US tax code, under a standard welfarist criterion, and under the horizontal equity adjusted criterion discussed in this paper. Finally, I provide a reference to the academic literature that I deem relevant. Under a standard Welfarist criterion, all tags are admissible as long as they provide information on any variable or function relevant to the planner's problem. Strikingly, the restrictions of horizontal equity prescribe a usage of tags that closely corresponds to reality. The classification of tags under horizontal equity is discussed below. The discussion is informal, in that many tags imply slightly different environments from the ones presented above.

My Yes-or-No classification is limited in the sense that I cannot actually check whether the criterion is violated by the actual tax code, but merely whether the tag that the tax code uses would be allowed under the criterion. For example, the blind or disabled may receive higher benefits or face lower taxes than those who are otherwise equal but not blind or disabled. In this respect, it is particularly relevant to consider what welfare levels the planner attributes to what individuals, since these are unobserved in practice. I restrict myself to observing that in such cases the tag does seem to be compensation for differences in welfare at given consumption and effort levels: it does not seem that there are many cases in which people attempt to blind or disable themselves because they would be better off as a blind or disabled person. Thus, the social planner's choices seem to provide them with more welfare than an otherwise equivalent blind or disabled person, despite the latter possibly receiving higher benefits.

3.1 Static Tags

Income is the classic basis for taxation, and can be seen as a tag for ability in the Mirrleesian setting. Consumption taxation is equivalent to income taxation in this setting, and Atkinson and Stiglitz (1976) show that Horizontal Equity (in a version somewhat different from mine) in combination with differences in tastes does not imply uniform taxation. Unemployment benefits are clearly intended for the involuntarily unemployed. Proof and reaffirmation of involuntariness are often required for unemployment benefits. This is a precise tag of earnings inability ($\theta = 0$), which our criterion admits regardless of the height of the benefit. Disability and blindness benefits can increase the consumption levels of the blind and disabled above those who have the same earnings ability and work the same hours, but are clearly also a tag of circumstance: those who are disabled or blind have a natural disadvantage in life, deriving lower welfare from the same economic circumstances. This fits the framework as differences in u^i , so that the tags are permissible.

Research on the role of household form and the number of children in taxation is ongoing, so that it is hard to draw definite conclusions on their implications for horizontal equity as defined here. When welfare is treated at the household level, then the welfare of a household may well be assigned different cardinality by the planner. Issues such as home production and household returns to scale may also affect welfare levels. Overall, it is plausible that tags for household composition are allowed under our criterion. Similar observations hold for the number of children. In addition, when treating both tags in a purely static sense, they may be seen as choices. Mortgage interest deductions, exemptions for charitable contributions, and exemptions for health expenditures are less discussed in the academic literature, but all part of the tax code. The former two clearly relate to choices. Therefore these tags do not violate the narrow horizontal equity restriction: any agent who behaves like another is treated the same. The latter should be seen similar to tags of blindness and disability: they are tags of the welfare derived from a given income.

Finally we arrive at a large number of tags that are not used in practice, and are clearly diffuse in the sense of this paper. These include gender, height, race, genetic information, and many others. Mankiw, Weinzierl, and Yagan (2009) additionally list the following: skin color, physical attractiveness, and parents' education.

3.2 Dynamic Tags

For dynamic tags, we use the adjusted criterion as above, where agents are compared to other agents of the same age on a per-period basis. As already discussed, while income and assets can be used, past income and past assets are excluded under horizontal equity. Age-based taxation is allowed, although I have already qualified earlier the extent to which this depends on how the criterion is carried into a dynamic setting.

Focusing on life-cycle comparisons instead of the criterion proposed here would address another aspect of tax systems that is sometimes considered unfair: volatile incomes are, under progressive tax systems, taxed more heavily than less volatile ones. This issue was first addressed by Vickrey (1939). On the other hand, this would rule out narrowly age-dependent schemes. In fact, while the US tax code does make taxes somewhat dependent on age, it does not do so too sharply: rather, it makes exceptions for the young (be it through supplementary systems such as schooling, federal student loans, etcetera) and the elderly (in the code as well as through retirement programs). The same goes for the tax code of many other countries. On the other hand, while the US tax code no longer provides for income averaging to smooth out the effect of volatile incomes (it did in the past), such provisions are provided in the tax codes of many other countries. Seeing all this, it seems that there is indeed concern for horizontal equity that allows for a distinction by age, albeit perhaps not

as sharply as a year-by-year scheme prescribes: income averaging, where it is permitted, is often very local (not over the entire life cycle but over a few consecutive years), while age-based taxation is applied to much longer phases of the life-cycle than by years. It appears that concern for horizontal equity is not quite over the entire life cycle, but also not quite limited to single years. In some sense, this is obvious: there is no inherent value to age as measured in years, we just want to compare *sufficiently similar* agents.

A tag on education would be diffuse as well, and therefore excluded. A large literature on dynamic optimal taxation with human capital typically considers human capital unobservable. But where it is considered observable, education is often pointed to as the observable entity. This paper suggests that we are unlikely to observe such policies in practice, while we do already see income-contingent student loans, which would be permissible under Horizontal Equity.

3.3 Shortcomings

At last, a note of nuance is in order: while the above analysis shows that the concept of Horizontal Equity as presented here can explain which tags are and are not used in the United States, there is no reason to expect this relation to always hold perfectly. There are differences in social norms across groups, countries, and over time. There are many examples of suppression and corruption that obviously violate horizontal equity, both in the past and today. And some countries' tax systems today feature democratically elected divergences from the empirical rule above, even if they are mostly small. For example, a set of South African policies that are collectively referred to as Black Economic Empowerment do use race as a tag. Here, my undertaking of positive optimal taxation fails.

4 Conclusion

As stated in the introduction, the absence of the use of tags is arguably the biggest difference between the practice and the theory of taxation today. I posit that the above is the most successful attempt so far at reconciling the two. On the way to achieving this, the paper has found an operationalization of Horizontal Equity in a utilitarian framework that does not suffer from the usual criticisms of the concept. In particular, it was never necessary to define a 'natural state' of any sort. Instead, the notion of *equal treatment* considered here is close to the way fundamental rights are defined in today's laws.

The notion of Horizontal Equity has been on a long slide in the field of economics. Where Rawls (1971) and Nozick (1974) are often presented as equal polars in undergraduate courses on the philosophy of economics, the practice of economic research on taxation clearly tends towards the former. Three things may be responsible for this: the difficulty to operationalize

notions of Horizontal Equity, the initial success of the strictly Welfarist setup of the original work by Mirrlees (1971), as well as the focus on Pareto optimality within the economics profession. The empirical success of this paper perhaps suggests that concepts like Horizontal Equity should perhaps be given a more central role when we try to understand the societies in which we live.

This paper is surely not the end of the road. As was evident in the above, we have no clear way of thinking about welfare heterogeneity (differences in utility functions) within Welfarist frameworks, despite their general acceptance in economics. Also, this paper has operationalized Horizontal Equity, but there may be other ways of doing this, and in particular in a dynamic context there is plenty of room for doubt. Research on positive optimal taxation is promising: because optimal taxation is perhaps the most normative field in all of economics, studying how taxes are raised in practice is a most promising avenue for understanding how social preferences arise.

References

- Akerlof, G. A. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American Economic Review*, 8–19.
- Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The Review of Economic Studies* 73(1), 1–30.
- Alesina, A., A. Ichino, and L. Karabarbounis (2011). Gender-based taxation and the division of family chores. *American Economic Journal: Economic Policy* 3(2), 1–40.
- Atkinson, A. B. and J. E. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics* 6(1), 55–75.
- Blumkin, T., Y. Margaliot, and E. Sadka (2009). Incorporating affirmative action into the welfare state. *Journal of Public Economics* 93(9), 1027–1035.
- Blumkin, T. and E. Sadka (2007). A case for taxing charitable donations. *Journal of Public Economics* 91(7), 1555–1564.
- Bohacek, R. and M. Kapicka (2008). Optimal human capital policies. *Journal of Monetary Economics* 55(1), 1–16.
- Diamond, P. A. and J. A. Mirrlees (1978). A model of social insurance with variable retirement. *Journal of Public Economics* 10(3), 295–336.
- Domeij, D. and P. Klein (2013). Should day care be subsidized? *The Review of Economic Studies* 80(2), 568–595.
- Farhi, E. and I. Werning (2013). Insurance and taxation over the life cycle. *The Review of Economic Studies* 80(2), 596–635.
- Fleurbaey, M. and F. Maniquet (2006). Fair income tax. *Review of Economic studies*, 55–83.
- Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies* 70(3), 569–587.
- Hendren, N. (2014). The Inequality Deflator: Interpersonal Comparisons without a Social Welfare Function.
- Immonen, R., R. Kanbur, M. Keen, and M. Tuomala (1998). Tagging and taxing: The optimal use of categorical and income information in designing tax/transfer schemes. *Economica* 65(258), 179–192.

- Jacquet, L. and E. Lehmann (2015). Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous.
- Kaplow, L. (1989). Horizontal Equity: Measures in Search of a Principle. *National Tax Journal*, 139–154.
- Kaplow, L. and S. Shavell (2001). Any non-welfarist method of policy assessment violates the pareto principle. *Journal of Political Economy* 109(2), 281–286.
- Kleven, H. J., C. T. Kreiner, and E. Saez (2009). The optimal income taxation of couples. *Econometrica* 77(2), 537–560.
- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation. *Econometrica* 73(5), 1587–1621.
- Logue, K. and J. Selmirod (2008). Genes as Tags: The Tax Implications of Widely Available Genetic Information. *National Tax Journal*, 843–863.
- Mankiw, N. G. and M. Weinzierl (2010). The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution. *American Economic Journal. Economic Policy* 2(1), 155.
- Mankiw, N. G., M. Weinzierl, and D. Yagan (2009). Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives* 23(4), 147–74.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 175–208.
- Nozick, R. (1974). *State, anarchy, and utopia*. Basic Books.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Saez, E. (2002). Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *Quarterly Journal of Economics* 117, 1039–1073.
- Sausgruber, R. and J.-R. Tyran (2014). Discriminatory taxes are unpopular Even when they are efficient and distributionally fair. *Journal of Economic Behavior & Organization* 108, 463–476.
- Stantcheva, S. (2015). Optimal taxation and human capital policies over the life cycle.
- Vickrey, W. (1939). Averaging of income for income-tax purposes. *The Journal of Political Economy*, 379–397.
- Weinzierl, M. (2011). The surprising power of age-dependent taxes. *The Review of Economic Studies* 78(4), 1490–1518.
- Weinzierl, M. (2014). The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *Journal of Public Economics* 118, 128–142.
- Werning, I. (2007). Pareto efficient income taxation. *Massachusetts Institute of Technology, Manuscript* 44, 45.