NBER WORKING PAPER SERIES

TAX KNOWLEDGE AND TAX MANIPULATION: A UNIFYING MODEL

Ashley C. Craig
Joel Slemrod

Working Paper 30151
http://www.nber.org/papers/w30151

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2022

We are grateful to Antoine Ferey, Nathan Hendren, Jim Hines, Louis Kaplow, Kai Konrad, Kory Kroft, John Leahy, Dylan Moore, Christian Moser, Alex Rees-Jones, and Nicolas Werquin for comments and suggestions, as well as seminar participants at LSE, Oxford, Princeton University, Tsinghua University, the Office of Tax Policy Research M-TAXI 2021 Conference, the Online Public Finance Seminar, the National Tax Association, and the Paris Tax Workshop. Financial support from the University of Michigan is gratefully acknowledged. There are no other sources of research support to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Ashley C. Craig and Joel Slemrod. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Tax Knowledge and Tax Manipulation: A Unifying Model
Ashley C. Craig and Joel Slemrod
NBER Working Paper No. 30151
June 2022
JEL No. H2,H21,H26

ABSTRACT

We provide a unified analysis of taxation and taxpayer education when individuals have an incomplete understanding of a complex tax system. The analysis is independent of whether income is earned legitimately, or by avoiding or evading taxes. In this sense, learning about tax minimization strategies (tax manipulation) is isomorphic to learning about tax rates. The government in our model balances a trade-off: A better understanding of the tax system potentially allows taxpayers to optimize more effectively, but also affects government revenue. Optimal taxpayer education and the optimal amount of redistribution can both be characterized by aggregate sufficient statistics, which do not require information about how biases or behavioral responses vary across the decision margins. We provide similarly simple rules for how tax rates on different income-generating activities should be set relative to each other.

Ashley C. Craig
Department of Economics
611 Tappan Ave
258 Lorch Hall
Ann Arbor, MI 41809
United States
ashcraig@umich.edu

Joel Slemrod
University of Michigan
701 Tappan Street
Room R5396
Ann Arbor, MI 48109-1234
and NBER
jslemrod@umich.edu

1 Introduction

Taxpayers face complex tax systems, which many struggle to understand while others strive to exploit. Survey and experimental evidence suggests pervasive misperceptions. A quarter of individuals do not know whether people with higher incomes owe a larger share of their earnings in income taxes, and more than 10 percent believe everyone owes the same share (NPR/Kaiser/Kennedy School, 2003). Taxpayers even misunderstand aspects of the tax system that affect their own choices: Experiments show that many taxpayers use the more easily calculated average tax rate as their marginal tax rate (Rees-Jones and Taubinsky, 2020), and that complex tax rules cause people to misunderstand the embedded incentives (Abeler and Jäger 2015).[1] Some people fail to take up the Earned Income Tax Credit (EITC) when they are eligible (Currie 2008, Bhargava and Manoli 2015); and consumers under-react to commodity taxes when they are not included in posted prices (Chetty, Looney, and Kroft, 2009).[2]

Misperceptions about taxes are not fixed. A store can post prominent tax alerts on their shelves. The federal government could mount a taxpayer education program that alerts likely qualifiers about the EITC, or patiently walks those wanting to be informed through the difference between an average tax rate and a marginal tax rate. Even without help, Morrison and Taubinsky (2019) show that consumers can and do use costly mental effort to increase the accuracy of their assessment of a tax-inclusive price when the stakes increase. Similarly, taxpayers could choose to acquire information from the internet (Hoopes et al. 2017) or purchase information from tax professionals.

These observations apply equally to taxpayers who seek to learn about the "dark arts" of evasion and abusive avoidance. Clearly, some people learn about the tax system to find ways around it: The IRS (2019) estimates the annual gross federal tax gap in tax years 2011-13 at $441 billion, or about 16.4% of what should have been remitted.[3] And the extent of evasion is mutable in the same way as knowledge individuals need to make fully compliant decisions. For example, the tax planning industry sells information about tax-reducing strategies that span legal and illegal activities.[4] Government policy also affects

[1] Confusion of average and marginal tax rates was dubbed "ironing" by Liebman and Zeckhauser (2004). See also experimental work by de Bartolme (1995).

[2] Chetty, Friedman, and Saez (2013) also present evidence that a lack of knowledge reduces the amount of bunching that is observed around kink points in EITC schedules. Bastani and Waldenstrom (2021) show that taking advantage of subtle tax incentives is strongly associated with cognitive skill, which further suggests that these differences in behavior are due to misperceptions.

[3] Recent evidence suggests that evasion among the rich may be much higher than is uncovered by the methods used in the IRS tax gap studies (Guyton et al. 2021).

[4] Knowledge may also spread between taxpayers. Battaglini et al. (2020) show that information is transmitted between clients served by the same tax professional. The results of Boning, Guyton, Hodge and

taxpayers' perceptions of the enforcement environment: Evasion is affected by a wide range of tax enforcement policies, ranging from the volume and targeting of audits to the penalties invoked for discovered evasion.[5] The fact that the IRS publicizes successful enforcement initiatives in the weeks before April 15 (Blank and Levin 2010) suggests that they want to make such enforcement more salient.[6]

Building on the similarity between knowledge about legitimate and illegitimate activities, we provide a unified normative analysis of the entire tax system. This includes the optimal amount of taxpayer knowledge, redistribution and how tax rates should be set relative to each other. The analysis is independent of whether income is earned legitimately or, alternatively, by avoiding or evading taxes. Regardless of the type of knowledge in question, the government strikes a balance between helping taxpayers optimize more effectively and raising revenue. Optimal government policies are always characterized by the same set of aggregate sufficient statistics. In this sense, the analysis of government policy toward these two aspects of taxpayer knowledge is isomorphic.

To gain insight into these issues, we abandon the idea that all individuals choose how much of a pre-determined set of tax parameters to absorb and, therefore, react to. In its place is a formal and very general depiction of the reality that characterizes all modern tax systems—that there is a vast assortment of both intended tax rules and tax-minimizing devices that savvy, willing taxpayers can take advantage of. In addition to the many legitimate ways in which individuals might respond as they learn more, they may also engage in all kinds of tax-reducing strategies. These could range from the relatively straightforward, such as deducting the costs of a home office, to the more complicated such as evasion using offshore accounts (see Alstadsæter, Johannesen, and Zucman 2019).

In our model, every individual has some perception of a potentially complex tax system. This includes the relevant income tax rate schedule, as well as such matters as the deductions and credits for which one qualifies, the tax benefits of reclassifying potential business expenses, and the expertise required to set up offshore tax avoidance or evasion schemes. The totality of these decisions determine both her income and her tax burden. However, the taxpayer fails to fully optimize because she is not completely aware of the tax implications of these activities.

We start by taking the tax schedule as given and asking whether the government should educate taxpayers, partially correcting their misperceptions. This choice boils

Slemrod (2020) suggest information transmission between firms who receive enforcement letters.

[5]Closing inconsistencies that facilitate "tax arbitrage" (Bradford 1980) reduces the extent of legal tax avoidance. Keen and Slemrod (2017) characterize optimal enforcement policy with these types of tools.

[6]Given the objectively low rate of audit coverage, it might be that moving people's perceptions away from the truth would decrease evasion and its attendant social costs. This is true in our model as well.

down to a trade-off between two forces. First, there is an internality: A better understanding of the tax system may allow a taxpayer to optimize more effectively—although we also highlight that learning can be counter-productive for the taxpayer in some cases. Second, any change in the taxpayer's income leads to a fiscal externality. If, for example, she works more, she ignores the fact that government revenue increases. If she evades more, she ignores the resulting decline in government revenue.

Assuming that a better understanding of the tax system is directly beneficial to the taxpayer being educated, it is also socially beneficial if society places a large enough welfare weight on that taxpayer. For such a taxpayer, the internality correction dominates concerns about government revenue. At the opposite extreme, if the taxpayer's welfare weight is very low, the fiscal externality dominates: Education is beneficial if and only if it increases government revenue. Away from these extremes, our sufficient statistics formulas show how to optimally balance the internality and fiscal externality.

These lessons change if the government is able to fine-tune taxpayers' perceptions of the tax system, while also adjusting the tax system itself. In that case, the government optimally creates the impression that all activities are treated equally by the tax system as long as different choices about how to earn a given level of income do not directly reveal anything about an individual's intrinsic type. This is true regardless of an individual's social welfare weight, and is independent of how the true tax system treats different ways of earning income. The intuition is that every taxpayer's effort is allocated most efficiently if perceptions are harmonized in this way, while a non-linear tax on overall income suffices to achieve any redistributive aims of the government.

Our optimal tax results mirror these lessons. The optimal amount of overall redistribution is pinned down by a condition similar to Mirrlees (1971), except that the planner balances redistribution with the internality. If types with low incomes have high marginal social welfare weights, the planner aims to reduce the internality for such taxpayers by minimizing the gap between perceived and actual tax rates. For types with high incomes, whose welfare weights are lower, the internality is less important. If welfare weights converge to zero at very high incomes, top tax rates are set to maximize revenue, just as in standard optimal taxation problems with perfect knowledge of the tax system.

We also provide guidance on how marginal tax rates should be set on different activities, again balancing the internality and fiscal externality. If all tax rates are equally "salient"—i.e, taxpayers pay equal attention on every margin—and types do not differ in their intrinsic preferences regarding how to earn their incomes, relative tax rates are set so that a weighted average of the perceived and actual tax rates is equalized across decisions. The weight on the perceived and actual tax rates depends on the value society places on taxpayers at each level of income at the margin. For taxpayer types with very low social welfare weight, actual tax rates are equalized, which maximizes government revenue. For types with very high welfare weights, internalities are equalized across decisions. For a taxpayer with the average marginal social welfare weight, *perceived* tax rates are equalized, which ensures overall efficiency in a taxpayer's decisions.

In all of these components of our analysis, we demonstrate that substantial progress can be made using expressions involving just a few aggregate sufficient statistics: average perceived and actual marginal tax rates, the overall response of income, and the marginal social welfare weights of the affected taxpayers. The same sufficient statistics formulas continue to apply in other versions of the model, for example with endogenous knowledge acquisition by taxpayers. Importantly, no information is required about how misperceptions vary across the many activities via which taxpayers earn their income. This opens up the possibility that future experimental or empirical work could quantify the welfare gains from knowledge provision in a relatively straightforward way.

Prior literature on optimal taxation with behavioral biases contains some discussion of policies that alter taxpayer misperceptions.[7] In Goldin (2015), the government chooses between high- and low-salience taxes on a particular good, trading off the lower excess burden of low-salience taxes against the optimization errors that reduce welfare because misperceiving consumers misallocate spending among consumption goods. Taubinsky and Rees-Jones (2018) address the welfare effects of changing the distribution of tax salience, while Rees-Jones and Taubinsky (2020) examine the welfare effects of increasing the fraction of taxpayers who use the ironing heuristic. Farhi and Gabaix (2020) develop a model of optimal nudging that addresses the optimal combination of nudges and commodity taxes in problems with both biased behavior and externalities.[8] Moore and Slemrod (2021) characterize optimal bias alteration policies.

In this literature, investments in knowledge about the tax system are directed at learning the truth, which informs the decisions of fully compliant and, implicitly, unaggressive taxpayers. But, while some taxpayers struggle to understand the tax system, others profit from tax complexity. We show that rigorously addressing the isomorphism between these

[7]Several studies show that information about other government programs changes behavior. An information letter documenting the kinked nature of the Norwegian disability insurance system altered behavioral responses and increased the earnings elasticity (Kostøl and Myhre, 2021). A brochure and web tutorial on key Social Security features increased the labor force participation of older workers (Liebman and Luttmer 2015). Bhargava and Manoli show that take-up of the EITC is sensitive to the frequency, salience, and simplicity with which information is provided to potentially eligible individuals.

[8]There is also a literature on optimal policy with behavioral agents in other contexts. For example, several studies consider optimal paternalistic savings policies, such as default or mandatory retirement contribution plans (Bernheim, Fradkin, and Popov 2015; Goldin and Reck 2021; Fadlon and Laibson 2017). Lockwood (2020) and Moser and Olea de Souza e Silva (2019) look at optimal policy if people have present bias.

two aspects of taxpayer knowledge yields important new insights regarding the optimal structure of tax systems—both tax rate schedules and the extent and distribution of knowledge. The fact that this has not been considered before is surprising given the intense policy attention focused on increasing inequality of income and wealth and the prevalence of tax evasion among the rich. For many members of this group, the issue is not to what extent they understand the tax system as it was intended, but rather to what extent they can learn to manipulate it to their advantage via unintended use of legal transactions or ex ante attractive gambles on illegal behavior in the form of evasion.

We view our unification of these strands of inquiry and our observations about the isomorphism of salience and manipulation as extending the insights of the literature on the elasticity of taxable income that was developed by Feldstein (1999), in which the "anatomy" of behavioral response is irrelevant for the welfare analysis of tax rate changes. Building on this literature, Keen and Slemrod (2017) unify the optimal tax systems analysis of tax rate and tax enforcement choices. Our analysis encompasses both of these insights, and extends the unification to include policies that rectify policy misperceptions and address taxpayer manipulation of the tax system for their private benefit.

2 An Individual's Perception of the Tax System

Consider a unit mass $\mathcal{I}$ of individuals. Individual $i \in \mathcal{I}$ makes a vector of decisions, $\boldsymbol{x}^i = \left[x_1^i, x_2^i, \ldots x_j^i \ldots, x_{|\mathcal{J}|}^i\right] \in \mathbb{R}_+^{|\mathcal{J}|}$ where $|\mathcal{J}|$ is the cardinality of the set of decisions, $\mathcal{J}$. Decisions could include such choices as labor supply, the extent to which she recharacterizes ordinary income as capital gains, and tax evasion. A function $y(\boldsymbol{x}^i)$ maps her decisions to her income, $y \in [0, \overline{y}]$. For simplicity, we assume that $y_j = \frac{\partial y(\boldsymbol{x}^i)}{\partial x_j} \geq 0$, and that y and x_j^i are defined on closed intervals.[9] Of her income, the individual consumes $c^i - T(\boldsymbol{x}^i)$ where $T \in \mathcal{T} \subseteq \mathcal{C}\left(\mathbb{R}_+^{|\mathcal{J}|}, \mathbb{R}\right)$ is a non-linear tax system.[10] Her disutility from choosing $\boldsymbol{x}^i$ is $\chi^i\left(\boldsymbol{x}^i \mid \theta^i\right)$, which is twice continuously differentiable.

We also assume for simplicity that utility is quasilinear in consumption. Relaxing this assumption has little impact on the results regarding knowledge provision except that an individual's decision then depends on their perceptions of their overall tax burden as well as their marginal tax rate.[11] However, it is a more substantive restriction when it comes to

[9] The assumption that y increases with x_j^i is not important but simplifies our discussion of optimal knowledge provision. However, monotonicity of $\partial y / \partial x_j$ substantially simplifies the analysis of optimal taxation.

[10] We use $\mathcal{C}\left(A, B\right)$ to denote the space of continuous functions mapping from A to B. In some parts of the paper, we restrict $T\left(z\right)$ to be twice continuously differentiable.

[11] The same sufficient statistics apply for optimal knowledge policy with a general utility function. However, the marginal value of public funds differs from one in that case because it must take into account the

the analysis of optimal taxation in Section 4, so we relax it in Section 4.3.

Bringing everything together, the true "experienced" utility of individual i is given by equation 1, and is a function of the true tax system, T. It is also a function of her type, θ^i.

$$U_i = c^i - \chi^i(\boldsymbol{x}^i \mid \theta^i) = y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i) \tag{1}$$

The individual knows her type and the income-generating function, $y(\boldsymbol{x}^i)$. However, her perception of her tax liability may differ from the truth. She perceives a tax burden of $\widetilde{T}(\boldsymbol{x}^i \mid k, T)$ which depends on the true tax system and the state of her knowledge, k.[12] Her "decision" utility is thus given by equation 2.

$$\widetilde{U}_i = y(\boldsymbol{x}^i) - \widetilde{T}(\boldsymbol{x}^i \mid k, T) - \chi(\boldsymbol{x}^i \mid \theta^i) \tag{2}$$

Note that the disutility function, $\chi(\boldsymbol{x}^i \mid \theta^i)$, varies with an individual's type, which may capture characteristics such as cognitive ability.

We assume that $\widetilde{T}(\boldsymbol{x}^i \mid k, T)$ and $T(\boldsymbol{x}^i)$ are twice continuously differentiable with respect to $\boldsymbol{x}^i$, and that $\widetilde{T}(\boldsymbol{x}^i \mid k, T)$ is continuously differentiable with respect to k. Taking the derivative of equation 2 yields a first order condition for decision x_j.

$$y^i_j(\boldsymbol{x}^i) - \widetilde{T}_j(\boldsymbol{x}^i \mid k, T) = \chi_j(\boldsymbol{x}^i \mid \theta^i) \; \forall j \in \mathcal{J} \tag{3}$$

where y^i_j and $\chi_j(\boldsymbol{x}^i \mid \theta^i)$ are the partial derivatives $y(\boldsymbol{x}^i)$ and $\chi(\boldsymbol{x} \mid \theta^i)$ with respect to x^i_j, respectively, and $\widetilde{T}_j(\boldsymbol{x}^i \mid k, T)$ is the derivative of $\widetilde{T}(\boldsymbol{x}^i \mid k, T)$ with respect to x^i_j.

We make the following assumption throughout the paper, which is standard in the optimal taxation literature and greatly simplifies the analysis of both optimal taxation and taxpayer education.

Assumption 1. *The first order conditions (equation 3) jointly characterize a unique optimal choice of $\boldsymbol{x}^i$ for all but a finite set of individuals.*

We discuss sufficient conditions for Assumption 1 in Appendix A.1. The key requirement is that any interior decision utility-maximizing choice is unique, which ensures that taxpayers respond smoothly to marginal changes in knowledge.[13] We note that corner solu-

fiscal externality and potential internalities from a universal lump sum transfer. In turn, this affects the interpretation of the marginal social welfare weight, ψ^i.

[12]We will assume for now that this individual has no direct control over her level of knowledge, but we consider below the case in which she does.

[13]This is generically true, as any given individual having multiple utility-maximizing choices is a knife-edge case. However, it is important to note that adjustments would be needed if there were discrete changes in decisions among more than a finite set of individuals.

tions are not a problem for our results, as it does not matter for any of the welfare analysis whether the first order condition for x_j^i holds if x_j^i does not change with knowledge.[14]

This set of first order conditions captures many possibilities. As an arbitrary example, let x_1 be labor income. The remaining decisions could involve all kinds of deductions, tax avoidance and evasion schemes, incorporation of enterprises and so on.[15] Changing any of these other decisions can change both actual and perceived effective marginal tax rate on labor income. For example, the effective marginal tax rate on labor income may be lower if an individual can shelter a larger share of her income from taxation. In turn, a taxpayer's ability to shelter income depends on policy choices. If the government systematically pursues non-compliant high-income taxpayers more or less than they do low-income taxpayers, this changes the payoffs to avoidance and evasion schemes at different levels of income. Alternatively, they might exert particular effort to audit claimants of the EITC, as some evidence suggests is true in the United States (IRS 2021, Table 17).

In this framework, there is no conceptual difference between a taxpayer learning about the marginal tax rates that apply to legitimate income sources and a taxpayer learning about the benefits and risks of avoidance or evasion opportunities. Indeed, it will become clear as the analysis proceeds that tax salience and tax manipulation are inextricably related, and in some dimensions isomorphic. As taxpayers become more knowledgeable about the parameters of the tax system, they find opportunities to engage in behavior that leads to a private gain in welfare. This could be a shift in labor supply toward the level that is optimal given their true marginal tax rate. But it could also be taking advantage of a new deduction, incorporating a business, or shifting assets to a financial institution in the Cayman Islands. Although there is a private gain to all these types of knowledge, they also each have distinct implications for government revenue. To the extent that government revenue falls, learning is socially costly. If revenue increases (e.g., if individuals realize that their marginal tax rate is lower than they previously believed, or the audit probability higher that they had thought), there is an associated social benefit.

2.1 Defining Tax Rates

There are many tax rate concepts of interest in this model. First, we define the marginal tax rate on x_j as the share of any increase in income due to an increase in x_j that is taxed away by the government.[16] The perceived marginal tax rate is defined similarly.

[14] It is straightforward to rule out corner solutions with assumptions on the tax schedule and utility.

[15] We treat all decisions as continuous, but some choice are discrete in reality. The analysis can be extended to formally incorporate discrete changes, but this adds little conceptual insight.

[16] The special case of pure evasion involves a change in an individual's tax burden without any change in their true before-tax income. MTR_j^i is ill-defined in that case because it approaches negative infinity.

$$\mathrm{MTR}_j^i = T_j(\boldsymbol{x}^i)/y_j(\boldsymbol{x}^i) \qquad\qquad \widetilde{\mathrm{MTR}}_j^i = \widetilde{T}_j(\boldsymbol{x}^i \mid k, T)/y_j(\boldsymbol{x}^i) \tag{4}$$

Note that, in general, MTR_j^i varies with decisions other than x_j. For example, the effective tax treatment of labor income would vary depending on actions that an individual takes to shelter part of her income. Moreover, even at the same level of income, MTR_j^i could vary across individuals who earn that income in different ways.

In addition to the tax rates on individual decisions, each individual's actual and perceived *average* marginal tax rate will play a particularly important role in the analysis. We define these as the share of the *overall* increase in income that is taxed away by the government when an individual responds to some small policy change of size η — which may be a change to the tax schedule or the provision of knowledge.

$$\mathrm{MTR}^i = \frac{\sum_{j=1}^N \mathrm{MTR}_j^i y_j \frac{dx_j^i}{d\eta}}{\sum_{j=1}^N y_j \frac{dx_j^i}{d\eta}} \qquad\qquad \widetilde{\mathrm{MTR}}^i = \frac{\sum_{j=1}^N \widetilde{\mathrm{MTR}}_j^i y_j \frac{dx_j^i}{d\eta}}{\sum_{j=1}^N y_j \frac{dx_j^i}{d\eta}} \tag{5}$$

These average marginal tax rates are convenient sufficient statistics that simply answer the question of how much tax an individual pays (or perceives that they will pay) per dollar that his or her income increases with re-optimization. Mathematically, they are weighted averages of the marginal tax rates on each of the individual decisions that change when η changes, with the weight on MTR_j^i corresponding to the share of the overall income increase that comes from x_j.

3 Optimal Taxpayer Education

If individuals lack knowledge about the tax system or misperceive it, one solution is to educate them. We suppose that the government can provide information to taxpayers, which we capture as an increase in the stock of knowledge, k. This captures many possibilities: The government could produce or otherwise provide tax preparation software for free; professional tax advice could be taxed or subsidized; or information could be communicated explicitly as part of an information campaign. These policies could be universal or targeted at particular groups of taxpayers. For example, an education campaign about the EITC would largely affect low-income families.

In providing information to taxpayers, the government chooses a level of knowledge k from a feasible range, $K = [\underline{k}, \overline{k}]$. For now, we assume that the government directly chooses how much knowledge to provide, but in Section 5 we consider the alternative

However, the welfare gains and other key expressions below remain well-defined.

case in which individuals make their own costly investments to acquire knowledge.

Choosing a level of knowledge k comes at a social cost of $C(k)$, which we assume to be differentiable. This captures many possibilities. One is that a higher level of knowledge is always more costly to sustain. Alternatively, there may be costs to adjusting knowledge away from the status quo if information is required to change perceptions. Finally, there may be asymmetric costs that are incurred specifically if the government misleads taxpayers by reducing k.[17] In fact, if more knowledge is viewed has having a direct civic benefit, the social marginal cost of knowledge provision could even be negative instead of positive. We adopt a welfarist perspective throughout our discussion, but this type of asymmetry could easily incorporate a direct moral objection to government deception.

The government chooses the level of knowledge to maximize social welfare, $\boldsymbol{\mathcal{W}}$, which is an average across individuals of their experienced utilities after they have been transformed by a social welfare function, $\mathcal{W}$. This social welfare function is increasing and concave. Fixing marginal tax rates for now, we begin by studying optimal taxpayer education. This is one part of the overall problem, which is to maximize social welfare by choosing both a level of knowledge and the tax system.

$$\max_{k \in K, T \in \mathcal{T}} \boldsymbol{\mathcal{W}}(k, T) = \int_i \mathcal{W}\left(y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i)\right) di \tag{6}$$

It does this subject to the constraint that it must raise enough revenue to cover the cost of taxpayer education, plus a fixed quantity of exogenous government expenditure.

$$\int_i T(\boldsymbol{x}^i) di - C(k) \leq G \tag{7}$$

A welfarist perspective is embedded here in two ways. First, we do not discount the utility of individuals who evade their tax liability, although this can easily be accommodated and we discuss that possibility below. Second, we assume that maximization of decision rather than experienced utility is a mistake due to limited knowledge, and that only experienced utility is therefore of direct welfare relevance.

In our welfarist framework, a negative social benefit of knowledge implies that it may be optimal to deliberately mislead taxpayers by obfuscating the tax system or making it harder to find information. We stress, however, that the government's cost of knowledge provision can easily incorporate an asymmetric aversion to this type of deception as part of $C(k)$. Alternatively, our analysis can be viewed as providing guidance on how a gov-

[17] This could capture a deontological cost of deception, or political economy concerns such as endogenous mistrust of the government. Alternatively, Boccanfuso and Ferey (2021) argue that misleading taxpayers could reduce equilibrium welfare if taxpayers are inattentive but center their beliefs on the true tax system.

ernment should allocate a finite budget to target different types of taxpayers and decisions with education campaigns or other strategies to correct misperceptions.

None of our results regarding taxpayer education relies on the tax schedule being set optimally. In fact, a substantial practical advantage of our sufficient statistics approach is that the government need not even be aware of the true tax system in the broad sense that includes all of its loopholes and evasion techniques: As we discuss below, policy analysis requires only that the average actual and perceived tax rates on overall income are known. However, we do study optimal taxation in Section 4, as well as policies that affect both the tax system and taxpayers' perceptions of it (Section 3.9).

3.1 Behavioral Responses to Education

We assume that, when knowledge is provided (k is increased), the marginal tax rate that individual i perceives on decision x_j moves weakly closer to the true marginal tax rate, all else equal. Mathematically, $\partial \widetilde{T}_j / \partial k \leq 0$ if and only if $\widetilde{T}_j > T_j$, and $\partial \widetilde{T}_j / \partial k \geq 0$ if and only if $\widetilde{T}_j < T_j$. The same is true in expectation in an explicit model of belief formation (which we study in Section 5), but our approach with a single deterministic perception simplifies the analysis and exposition while preserving the economic intuition.

In response to the receipt of knowledge, taxpayers change their decisions. Specifically, differentiating i's first order conditions with respect to each x_j^i yields the Hessian matrix of $\widetilde{U}^i$, $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$. We assume that $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is invertible and that $\boldsymbol{x}^i$ is the unique utility maximizing choice (Assumption 1).[18] The implicit function theorem then applies, and the change in i's decision vector $\boldsymbol{x}^i$ can be written as:

$$d\boldsymbol{x}^i = \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times d\widetilde{T}^i \tag{8}$$

where $d\widetilde{T}^i$ is a vector of direct impacts on the marginal tax rate on each decision when knowledge is increased (i.e., $\partial \widetilde{T}_j / \partial k \times dk$).

3.2 The Internality from Education

There are two first order impacts of knowledge provision in addition to the mechanical cost of providing it, $C'(k)$. The first is that knowledge helps individuals re-optimize, moving closer (at least in simple cases) to the choice of $\boldsymbol{x}^i$ that maximizes their experienced utility. This manifests as an internality: When x_j^i changes, there is an impact on

[18] Invertibility ensures $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is negative definite, which slightly strengthens Assumption 1 so that the implicit function theorem applies. This is similar to the usual assumption in the optimal taxation literature that the second order condition holds strictly (Saez 2001). We discuss this further in Appendix A.1.

i's experienced utility which i does not anticipate. This impact is proportional to the gap between the perceived and true marginal tax rates on x^i_j, as that is the wedge between the marginal benefit and cost of engaging in that activity.

Aggregating across the various decisions, the overall impact on social welfare of individual i's re-optimization in response to increasing knowledge slightly is:

$$\textbf{Internality (IN}_i\textbf{)} = \psi^i \sum_{j \in \mathcal{J}} \left[\widetilde{\text{MTR}}^i_j - \text{MTR}^i_j \right] y_j(\boldsymbol{x}^i) \frac{dx^i_j}{dk} \tag{9}$$

where $\psi^i = \mathcal{W}'(U^i)/\lambda$ is the taxpayer's marginal social welfare weight relative to the value of a government dollar (λ). We assume that λ is one for the purpose of our examples throughout the paper, which implies that the average welfare weight is one.[19] Although ψ^i is endogenously determined as experienced utilities change, it could equivalently be treated as fixed for marginal policy changes. However, the distinction between fixed and endogenous weights does matter for larger policy changes. We also note that ψ^i could incorporate broader, non-welfarist, moral principles.

Because we assume that education weakly narrows the gap between the perceived and true marginal tax rates for each decision, one might be tempted to assume that more information is always beneficial. This need not be the case, however, because information that improves decision-making on one dimension may exacerbate distortions on a different dimension. For example, consider an individual who discovers a new way to shelter a fraction of her labor income. In response, she will shelter more of her income and be directly better off. However, she may also work and earn more: If she also underestimates her labor income tax rate, this would itself harm her welfare. On net, she may therefore be worse off or better off.

Nonetheless, there are several benchmark cases in which knowledge is always beneficial to its recipient. Proposition 1 provides three of these cases. Proofs of this and other technical statements are available in Appendix B.

Proposition 1. *An increase in k never lowers experienced utility if: (i) $\boldsymbol{x}^i$ is a scalar; (ii) utility is separable in each decision; or (iii) if $\widetilde{MTR}^i_j - MTR^i_j$ is independent of j for any value of k.*

The first of the three conditions is straightforward: If there is only one distortion, reducing it is beneficial.[20] The second condition follows directly from the same logic because

[19]This is true as long as a universal transfer does not cause individuals to re-assess their marginal tax rates. Otherwise, λ and the average welfare weight would differ from one.

[20]If $\widetilde{\text{MTR}}_j > \text{MTR}_j$, education narrows the gap by lowering $\widetilde{\text{MTR}}_j$ and x^i_j increases. If $\widetilde{\text{MTR}}_j < \text{MTR}_j$, then $\widetilde{\text{MTR}}_j$ rises and x^i_j falls. Either way, the resulting internality is positive.

education reduces the distortion on each decision margin, and those decisions do not interact. The third condition is more subtle, but works because it ensures that distortions are equally sized and that knowledge proportionally reduces all of them.

As an aside, note that Proposition 1 presumes that individuals trust the information they receive when the government educates them. This underlies our assumption throughout the paper that the direct effect of knowledge is to weakly reduce the gap between perceived and actual marginal tax rates on each decision. A deviation from this assumption—which may be plausible if the government has misled them in the past—would provide another reason why knowledge could be privately harmful to its recipient. While our sufficient results below would unaffected, this is important to bear in mind in thinking about which cases are empirically relevant.

3.3 The Fiscal Externality from Education

The second impact of taxpayer education is a fiscal externality, which arises because individuals ignore the impact of their decisions on government revenue. On each decision margin, the fiscal cost or benefit from i's re-optimization is proportional to the marginal tax rate on that decision.

$$\textbf{Fiscal Externality (FE}_i\textbf{)} = \sum_{j \in \mathcal{J}} \mathrm{MTR}_j y_j(\boldsymbol{x}^i) \frac{dx_j^i}{dk} \tag{10}$$

For example, suppose that an individual works more when informed that the marginal tax rate on labor income is lower than they thought. Not only does this lead to an impact on experienced utility for that individual (the internality), but it also raises government revenue. By contrast, informing a taxpayer about an evasion opportunity likely leads to a privately beneficial increase in evasion but a socially costly reduction in revenue.

3.4 Putting it Together: The Net Benefit of Education

Measured in government dollars, the total effect on social welfare when individual i responds to an increase in knowledge is the sum of the fiscal externality and the internality.

$$\Delta \boldsymbol{\mathcal{W}}_i = \underbrace{\sum_{j \in \mathcal{J}} \mathrm{MTR}_j y_j(\boldsymbol{x}^i) \frac{dx_j^i}{dk}}_{\textbf{Fiscal Externality}} + \underbrace{\psi^i \sum_{j \in \mathcal{J}} \left[\widetilde{\mathrm{MTR}}_j - \mathrm{MTR}_j \right] y_j(\boldsymbol{x}^i) \frac{dx_j^i}{dk}}_{\textbf{Internality}} \tag{11}$$

Directly using Equation 11 to assess the welfare impact of knowledge would require knowledge of both the response of each activity and misperceptions of the return to it.

This would be a formidable challenge, if only because decisions such as tax avoidance or evasion can typically not be directly observed. In fact, the government may not even know all of the details of how individuals avoid taxation of their incomes.

Equation 11 can, however, be rewritten in a simple way that involves only aggregate sufficient statistics. The overall gain in welfare is a weighted average of the actual and perceived fiscal externalities that occur with re-optimization:

$$\Delta\boldsymbol{\mathcal{W}}_i = \underbrace{(1-\psi^i)\text{MTR}^i \frac{dy^i}{dk}}_{\substack{\textbf{Actual Fiscal}\\\textbf{Externality}}} + \underbrace{\psi^i \widetilde{\text{MTR}}^i \frac{dy^i}{dk}}_{\substack{\textbf{Perceived Fiscal}\\\textbf{Externality}}} \tag{12}$$

where dy^i/dk is the overall response of i's income, and MTR^i and $\widetilde{\text{MTR}}^i$ are the actual and perceived average tax rates on income earned by i at the margin.

Equation 12 shows that just four objects are sufficient to assess whether greater knowledge is beneficial, aside from the direct cost of its provision: the overall response of income; the actual and perceived average marginal tax rates; and the social welfare weight placed on the recipient of knowledge.[21] Importantly, there is no need to know how perceived marginal tax rates differ across different activities, or even the composition of a taxpayer's income response.[22] The government does not even need to know precisely how individuals manage to avoide or evade taxation.

If the social welfare weight of the taxpayer being educated is sufficiently low at the margin, then the benefit of knowledge reduces to the fiscal externality. The government would want to educate such taxpayers if and only if doing so causes their decisions to change such that their overall tax burden increases by enough to offset any direct cost of educating them. It would never do so if it created a net drain on government funds. Two reasons stand out for why welfare weights might be very low for some recipients. First, ψ^i might be low for a taxpayers with very high incomes. Alternatively, society might steeply discount the marginal social welfare of an individual who evades taxes.

As we consider taxpayers with higher welfare weight, the perceived fiscal externality increases in importance until it reaches another important benchmark: when $\psi^i = 1$. Any benefit or cost to the affected taxpayer is then valued on par with a government dollar. In this case, only overall efficiency matters, which—as equation 12 highlights—is a matter of perceptions. This highlights an important point that we build on below: an individual's decisions are not distorted by actual taxation but rather by perceived taxation. By contrast, true redistribution occurs via the actual tax system. Thus, for example, redistributive

[21] This may be a subjective evaluation, or based on objective trade-offs as suggested by Hendren (2019).

[22] The main caveat is that MTR^i and $\widetilde{\text{MTR}}^i$ are specific to an individual and to a particular knowledge policy, as the composition of the response may vary across different taxpayers and reforms.

aspects of the tax system that taxpayers are unaware of are highly efficient because they achieve that redistribution without any corresponding distortion.

The final case is that of individuals with higher-than-average welfare weight ($\psi^i > 1$). The internality benefit to the recipient is then more important than government revenue at the margin. In the extreme, if ψ^i is large enough, equation 12 approaches being proportional to the gap between the perceived and actual fiscal externalities, $\widetilde{\mathbf{FE}} - \mathbf{FE}$. The government should then do everything it can to ensure that this taxpayer is well-informed (providing that knowledge is actually privately beneficial), even if the resulting behavioral responses cause a large fiscal cost.

3.5 The Optimal Level of Knowledge

If the level of knowledge is optimal, then the overall change in welfare – aggregated over all individuals, and accounting for the direct cost of its provision – must be equal to zero. Otherwise, there would be a first order gain to either providing more or less knowledge. This is summarized in Proposition 2, which provides a necessary condition for the level of knowledge to be optimal.

Proposition 2. *Consider an arbitrarily small perturbation that raises the level of knowledge. Given Assumption 1, the effect on social welfare is:*

$$\Delta \mathcal{W} = \underbrace{\int_i \left[(1-\psi^i) MTR^i + \psi^i \widetilde{MTR}^i \right] \frac{dy^i}{dk} di}_{\textbf{\textit{Benefit of Knowledge}}} - \underbrace{C'(k)}_{\substack{\textbf{\textit{Direct}} \\ \textbf{\textit{Cost}}}} \tag{13}$$

If the level of knowledge is optimal and $C'(k)$ is continuous, then equation 13 is zero.

3.6 Sufficient Statistics: Is Knowledge Beneficial?

Equation 13 can also be used to assess whether a taxpayer education campaign is welfare-improving at the margin. Here, we ask whether the aggregate benefit of knowledge is positive or negative. We focus on the case when the average marginal tax rate on income is positive, and is perceived to be positive as well.[23]

First, suppose that taxpayer i perceives that her tax burden would increase by more than it really would if she were to increase her income ($\widetilde{\text{MTR}}^i > \text{MTR}^i$). In this case, equation 13 says that taxpayer education is socially beneficial if and only if income rises with knowledge. This turns out to correspond with knowledge being privately beneficial

[23] This is for expositional simplicity, since it is the natural case. Analogous results apply with negative actual or perceived average marginal tax rates.

to its recipient, which includes any of the cases in Proposition 1. For example, suppose the taxpayer's only choice is her overall level of income. Then $\widetilde{\text{MTR}}^i$ would fall toward MTR^i as k rises, and y^i would always rise. Because education benefits its recipient and also raises government revenue, it is socially beneficial. By contrast, with many related decisions, it is possible that y^i falls. In that case, the recipient's welfare is directly harmed by a marginal increase in her knowledge. Since the fiscal externality is then also negative, taxpayer education is unambiguously socially harmful. In summary, if $\widetilde{\text{MTR}}^i > \text{MTR}^i > 0$, knowledge is socially beneficial if and only if it is privately beneficial.

Second, consider a taxpayer for whom $\widetilde{\text{MTR}}^i < \text{MTR}^i$, and suppose that knowledge is privately beneficial. Similar to before, this implies that the taxpayer's income falls with knowledge. This case is more complex to analyze because there is a trade-off between a positive internality and a negative fiscal externality. Providing knowledge is socially beneficial if and only if:

$$\psi^i > \hat{\psi} = \frac{\text{MTR}^i}{\text{MTR}^i - \widetilde{\text{MTR}}^i} > 1 \tag{14}$$

Equation 14 implies that knowledge should be provided if its recipient has a high welfare weight at the margin—at least greater than the population average of one—but not otherwise. If ψ^i is very low, the internality is relatively unimportant, and the negative fiscal externality implies that education lowers social welfare. But as we consider taxpayers with higher welfare weights, the positive internality may become important enough to outweigh the loss in government revenue.

The final case is another in which the taxpayer's experienced utility falls when she is educated because misperception on multiple dimensions causes her re-optimization to be counterproductive. Imagine that $\widetilde{\text{MTR}}^i < \text{MTR}^i$, but that income rises with knowledge. Then knowledge is socially beneficial in the opposite cases from above. Specifically, welfare increases if:

$$\psi^i < \hat{\psi} = \frac{\text{MTR}^i}{\text{MTR}^i - \widetilde{\text{MTR}}^i} \tag{15}$$

Here, the internality and fiscal externality again have opposite signs, but both are flipped: knowledge is privately harmful, but government revenue rises. The planner should provide knowledge only if ψ^i is low, so that direct harm to the recipient is discounted.

3.7 Examples

We now discuss three examples to illustrate how the analysis above applies to issues in the tax salience and tax manipulation literatures.

Example 1 imagines that taxpayers are confused and respond to their average tax rate instead of their true marginal tax rate. This is referred to as "ironing" in the literature (Liebman and Zeckhauser, 2004). Because marginal tax rates are higher than average rates in a uniformly progressive tax system, partially correcting this misperception would lead taxpayers to perceive their marginal tax rates as higher than otherwise. They would therefore reduce their incomes, causing government revenue to fall. But correcting their mistake also raises welfare directly via the internality. This trade off means that it will be optimal to correct a taxpayer's misperception if society places a relatively high social welfare weight on them at the margin, but not otherwise. This may correspond to informing low-income but not high-income taxpayers, although welfare weights could also capture other ethical criteria.

Example 1. *Let $\boldsymbol{x}^i$ be uni-dimensional and assume $MTR^i > \widetilde{MTR}^i > 0$ as would be true with the ironing heuristic and a progressive tax schedule. Then the change in x_1^i (equation 8) reduces to:*

$$\frac{dx_1^i}{dk} = \frac{\partial \widetilde{T}_1(\boldsymbol{x}^i, T)/\partial k}{y_{11}(\boldsymbol{x}^i) - \widetilde{T}_{11}(\boldsymbol{x}^i \mid k, T) - \chi_{11}(\boldsymbol{x}^i)} \tag{16}$$

where y_{11}, $\widetilde{T}_{11}$ and χ_{11} are the second derivatives of their respective functions with respect to x_1^i.

We know that $\partial \widetilde{T}(\boldsymbol{x}^i, T)_1/\partial k$ is positive, because $\widetilde{MTR}^i < MTR^i$; and the second order condition in the denominator is negative. Thus, $dx_1^i/dk < 0$ and $dy^i/dk < 0$. In turn, this implies that knowledge provision to this taxpayer is beneficial if and only if:

$$\psi^i > \hat{\psi}^1 = \frac{MTR_1^i}{MTR_1^i - \widetilde{MTR}_1^i} > 1 \tag{17}$$

This holds if the welfare weight is greater than a threshold $\hat{\psi}^1$ that exceeds the average of one.

Example 2 considers the net benefit of informing taxpayers about a tax avoidance opportunity. Despite the fact that tax avoidance is widely perceived as always being "bad", this example shows that it can be beneficial. In fact, precisely the same logic applies as in the example with ironing above: If the planner places sufficient welfare weight on the taxpayer who benefits from knowledge, then educating that taxpayer can produce a benefit that outweighs any loss in government revenue.

Example 2. *Let $\boldsymbol{x}^i$ be comprised of labor supply (x_1^i) and tax avoidance (x_2^i). Assume that the individual correctly perceives her marginal labor income tax rate ($T_1^i = \widetilde{T}_1^i$) but engages in less tax avoidance than is privately optimal because she is not aware of the benefits ($T_2^i < \widetilde{T}_2^i < 0$).*

Then x_2 increases, and knowledge provision to this taxpayer is beneficial if and only if:

$$\psi^i > \hat{\psi}^2 = \frac{\boldsymbol{-FE}}{\boldsymbol{IN}} = \frac{-\sum_{j\in\mathcal{J}} MTR_j^i y_j(\boldsymbol{x}^i)\frac{dx_j}{dk}}{\left[\widetilde{MTR}_2^i - MTR_2^i\right] y_2(\boldsymbol{x}^i)\frac{dx_2}{dk}} \tag{18}$$

Because x_2 increases and $T_2^i < \widetilde{T}_2^i < 0$, the denominator is positive (see Appendix B for a proof). Thus, the sign of $\hat{\psi}^2$ is the opposite of that of the fiscal externality.

If informing the individual of the benefits of tax avoidance lowers government revenue, the planner should inform an individual if and only if their welfare weight is high enough. But it is also possible that knowledge provision raises revenue if labor supply increases enough; in this case, there is no trade-off, and knowledge is always beneficial.

A second lesson of example 2 is that providing information about avoidance opportunities could actually increase revenue due to a type of "Laffer" effect. Although the direct effect of information provision is that the recipient avoids more tax, she may also supply more labor if avoidance lowers the effective marginal tax rate on labor income. This could more than offset the direct loss in revenue. Because both the internality and fiscal externality are positive, in this case taxpayer education is again unambiguously beneficial—regardless of the recipient's marginal social welfare weight.

3.8 Improvements in the Allocation of Effort

The benefit of knowledge embodied in equation 12 comes not only from the taxpayer adjusting their overall level of effort, but also from improvements in how they allocate effort across income-generating activities.[24] To see this most clearly, consider the case in which $\psi^i = 1$. As we highlighted above, this means that the planner cares only about overall efficiency, and the welfare benefit of knowledge is just the *perceived* fiscal externality.

$$\textbf{Perceived FE} = \sum_{j\in\mathcal{J}} \widetilde{\text{MTR}}_j y_j \frac{dx_j^i}{dk} \tag{19}$$

Further suppose for now that the taxpayer who is being educated does not change her overall level of income, $y(\boldsymbol{x}^i)$, but does change *how* she earns it. When is this reallocation of effort favorable from an efficiency standpoint? If the tax system were perceived by the taxpayer to treat all decisions symmetrically so that $\widetilde{\text{MTR}}_j$ is equalized across all decisions, there would be no welfare impact of such a compositional reallocation of effort.

[24]We abstract here from classical externalities that vary across different income earning activities, which also distort effort allocation (see Lockwood et al., 2017).

In general, however, welfare rises if she shifts her effort toward activities where the perceived marginal tax rate is *higher*.[25] Examining equation 19, this implies that dx^i_j/dk for activities where $\widetilde{\text{MTR}}_j$ is higher than average, so that the summation is positive.

The welfare gain or loss here stems from the fact that divergence among the perceived tax treatments of different income sources causes taxpayers to allocate their existing effort inefficiently. If the marginal tax rate on decision A is perceived to be higher than that on decision B, the taxpayer's effort is distorted toward B and away from A. Thus, a marginal reallocation of effort back toward A is efficiency-enhancing. For example, if the perceived tax treatment of production in Ohio is higher than that in Michigan, production is distorted toward Michigan. Thus, if knowledge causes a reallocation of production back toward Ohio, that improves efficiency, all else equal.

Example 3 reinforces these insights. The simplifying assumption is that taxpayers correctly perceive the simple average of marginal tax rates, but perceive marginal tax rates as differing more or less than they do in reality. For example, they might know the level of marginal tax rates in general but under-appreciate how they differ across activities. In this setting, knowledge is efficiency-enhancing if and only if the true tax system treats different activities more rather than less symmetrically than perceived.

Example 3. *Suppose the simple average of actual and perceived marginal tax rates is the same:*

$$\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} MTR^i_j = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} \widetilde{MTR}^i_j = \overline{MTR}^i$$

However, let one be more disperse: $MTR^i_j - \overline{MTR}^i = d[\widetilde{MTR}^i_j - \overline{MTR}^i]$ *for some constant* d*. Finally, assume knowledge reduces each wedge,* $\omega_j(k) = \widetilde{MTR}_j - MTR_j$*, by the same proportion. If* $\psi^i = 1$*, so that benefits to the individual and the government are equally weighted, the welfare gain is comprised of:*

1. *a fiscal externality that is directly proportional to the change in* i*'s overall income, and which is zero if* i*'s income does not change; and*

2. *the impact on welfare of re-allocation of effort across activities, which is positive if* $d < 1$*, negative if* $d > 1$*, and zero if* $d = 1$*.*

See Appendix B for a proof.

To understand the benefit or cost from re-allocation in this example, first observe that if $d > 1$, the true tax system treats different activities *more differently* than perceived. When

[25] A modified rule holds if $\psi^i \neq 1$, again involving a weighted average of perceived and true tax rates.

individuals learn this, their allocation of effort is further distorted. By contrast, if $d < 1$, learning more convinces them that the tax system treats the various income-earning activities *more equally* than they thought. Their allocation of effort improves. We build on this idea further in Section 3.9, where we show that the government can in some cases achieve welfare gains (even in the Pareto sense) by removing perceived differences in how different income-generating activities are taxed.

3.9 A Stronger Form of Knowledge Provision

To conclude our analysis of optimal knowledge provision, we consider a more extreme thought experiment in which the government has a very granular ability to affect individuals' perceptions of the tax system. Specifically, we suppose that the government can costlessly choose both the actual and perceived tax systems. We do this as a pedagogical exercise because the benefits of reducing distortions between different income sources are particularly clear. We do not propose it as a practical suggestion, since such precise and costless control of perceptions is clearly not possible in practice.[26]

As a benchmark, we also make a separability assumption about utility functions (Assumption 2). We return to what happens without this assumption below.

Assumption 2. *Individuals have experienced and decision utility of the following form:*

$$U^i = y\left(\boldsymbol{x}^i\right) - T\left(\boldsymbol{x}^i\right) - \chi\left(\boldsymbol{x}^i\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^i\right) \mid \theta^i\right) \tag{20}$$

$$\widetilde{U}^i = y\left(\boldsymbol{x}^i\right) - \widetilde{T}\left(\boldsymbol{x}^i\right) - \chi\left(\boldsymbol{x}^i\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^i\right) \mid \theta^i\right) \tag{21}$$

so that individuals do not differ in their preferences over $\boldsymbol{x}^i$, conditional on $y(\boldsymbol{x}^i)$.

This assumption states that types may differ in their disutility of earning a given level of income, but that any individual would choose to earn their income in the same way if they had the same level of income. As a result, the composition of effort chosen by an individual ($\boldsymbol{x}^i$) does not reveal anything about her that is not already revealed by the level of her income itself, $y(\boldsymbol{x}^i)$. This statement is formalized in Lemma 1.

Lemma 1. *The set of utility-maximizing choices of $\boldsymbol{x}^i$ given y^i are the same for all i.*

Building on this result, we can show that a Pareto gain can be obtained by harmonizing individuals' misperceptions about how different actions are treated by the tax schedule.

[26] If the planner could costlessly change perceptions in this way, it could completely obfuscate incentives created by the tax system by making taxpayers perceive it as entirely non-distortionary. It could then fully redistribute with no efficiency costs, equalizing experienced utilities across types. This related to Goldin (2015), who shows how to achieve it with two taxes on the same commodity which differ in their salience.

Specifically, the existing tax schedule can be replaced with one that achieves the same amount of redistribution across types but does not distort *how* each type earns a given amount of income. Because effort is allocated more efficiently, all individuals can be made better off providing that we also make appropriate compensatory transfers.

Proposition 3. *Suppose perceived marginal tax rates are not equalized across decisions for some: $\widetilde{T}_j(\boldsymbol{x}^i) \neq \widetilde{T}_j(\boldsymbol{x}^i)$. Given Assumption 2, there is a decision vector for each individual, a true tax system $T^\dagger(y(\boldsymbol{x}))$ and a harmonized perceived tax system $\widetilde{T}^\dagger(y(\boldsymbol{x}))$ such that:*

(i) Every individual is maximizing their utility given their perception of the tax schedule.

(ii) All individuals have strictly higher actual and perceived utility than before harmonization of the perceived tax system.

(iii) Perceived and actual government revenue are both weakly higher.

Proposition 3 shows that the government can make individuals of every type better off by ensuring that their beliefs are *harmonized* in the sense that they perceive the tax system as not affecting how they earn a given level of income. The intuition is that effort is better allocated for every type if any such distortions are removed.[27] Thus, each and every individual can produce the same income in a less costly way. This is a strong conclusion: Not only does this change produce a Pareto gain, but there is again no need to assume that the tax schedule is set optimally—either before or after the reform. Moreover, the same logic applies whenever we are considering informing an individual of a potential tax benefit, regardless of whether that involves "legitimate" activities or evasion.

This is a different conclusion from Section 3.6. Take Example 2, where we held the tax schedule fixed and considered the benefit of education. It was beneficial to inform a high welfare weight taxpayer of a pure avoidance opportunity. This obtained because the private gain from re-optimization outweighed any loss in government revenue. Here, where we allow a simultaneous adjustment to the tax schedule, it would be better *not* to inform taxpayers of such an opportunity. Instead, we can avoid the resulting distortion to their effort, and just lower their type's overall tax burden. The reason for the difference in results is that the government's more granular ability to affect perceptions allows it to separate efficiency in the allocation of effort from redistribution.

Without separability (Assumption 2), we would not have obtained as strong a conclusion. The reason is that different types would then have different preferences regarding

[27] Since non-linear taxation is available, there is no gain from having different perceived tax rates on goods that are more or less elastic, and Assumption 2 rules out any "tagging" benefit to differential taxation that would arise if different types had different preferences about how to earn a given income.

how to earn their income. We would therefore have an incentive to have higher perceived marginal tax rates on activities that are favored by high-ability types (or more generally, types with lower welfare weight).[28] For example, it may be optimal to impose harsh penalties for tax evasion if higher types are *intrinsically* more likely to tax evade—in the sense that they would evade taxes more than lower types even if they had the *same* level of income. Note that this is a statement about intrinsic preferences, which is different from higher types simply evading more than lower types in equilibrium.

4 Optimal Taxation

We now characterize the optimal tax rate schedule, holding fixed the level of knowledge. As before, we emphasize that the results here apply equally to choices between legitimate income sources, and to tax minimizing activities. Taxpayers' imperfect knowledge of the tax system has two consequences. First, they may not respond to changes in the tax system, either because they are not fully aware of them or because they make mistakes in understanding such reforms. Second, an internality arises, which is again proportional to the gap between the actual and perceived average marginal tax rates.

We follow a perturbation approach similar to Gerritsen (2016a). We start from an initial tax schedule, $\mathcal{T}\left(\boldsymbol{x}^i\right)$. Then we consider a reform that is characterized by a direction, $\tau\left(\boldsymbol{x}^i\right)$, and a size, η. This allows us to parameterize the post-reform tax schedule, $T\left(\boldsymbol{x}^i,\eta\right)$, as follows.

$$T\left(\boldsymbol{x}^i,\eta\right)=\mathcal{T}\left(\boldsymbol{x}^i\right)+\eta\tau\left(\boldsymbol{x}^i\right) \tag{22}$$

Using this notation, we can write experienced utility as a function of the parameterized tax schedule:

$$U^i=y\left(\boldsymbol{x}^i\right)-T\left(\boldsymbol{x}^i,\eta\right)-\chi\left(\boldsymbol{x}^i\mid\theta^i\right) \tag{23}$$

and we can similarly write decision utility as a function of the reform:

$$\widetilde{U}^i=y\left(\boldsymbol{x}^i\right)-\widetilde{T}\left(\boldsymbol{x}^i,k,\eta\right)-\chi\left(\boldsymbol{x}^i\mid\theta^i\right) \tag{24}$$

From here, we evaluate a marginal reform by considering a small change, $d\eta$. Then we ask whether there is a welfare gain from any such small change in the tax schedule. This optimization over the set of potential tax systems is essentially equivalent to an informational constraint, which is that the social planner can observe $\boldsymbol{x}$ but not θ^i.

[28]Starting with Akerlof (1978), there is a rich literature on using innate differences in preferences to tag higher types and relax the incentive constraints in optimal tax problems.

The analysis is complicated by the multi-dimensional nature of individual decisions. Nonetheless, we are able to provide substantial insight by adopting assumptions that are restrictive but which resemble those elsewhere in the optimal taxation literature. For simplicity, we start by assuming that individuals' income-earning choices ($\boldsymbol{x}^i$) respond to changes in their marginal tax rates but not to level of their tax burden.

Assumption 3. *The perceived marginal tax rate on decision j, $\widetilde{T}_j\left(\boldsymbol{x}^i, k, \eta\right)$, is a smooth function only of k, $\boldsymbol{x}^i$, and the true marginal tax rate on that decision, $T_j\left(\boldsymbol{x}^i, \eta\right)$.*

We relax this assumption in Section 4.3. For now, we note that ruling out "income effects" in labor supply is common in the optimal tax literature, but is more restrictive here. For example, it rules out "ironing", because ironing involves individuals responding to their *average* tax rate, which is a function of the level of their tax burden.

In addition to Assumptions 1 and 3, we also adopt two further assumptions to ensure tractability of our analysis of optimal taxation.

Assumption 4. *Decision utility has increasing differences in $\boldsymbol{x}$ and θ_i. Specifically:*

- *The net decision utility gain from each decision x_j increases with each other decision, $x_{j'}$.*

$$y_{jj'}(\boldsymbol{x}^i) - \widetilde{T}_{jj'}(\boldsymbol{x}^i, k, \eta) - \chi_{jj'}(\boldsymbol{x}^i \mid \theta^i) \geq 0 \tag{25}$$

- *The net decision utility gain from each decision x_j increases with an individual's type, θ^i: i.e., $\partial\chi_j(\boldsymbol{x}^i \mid \theta^i)/\partial\theta^i < 0$ for all i and $\boldsymbol{x}$.*

Assumption 5. *Higher types experience lower total disutility given the same decision vector, $\boldsymbol{x}$: for all $\boldsymbol{x}$, $\chi(\boldsymbol{x} \mid \theta^i) < \chi(\boldsymbol{x} \mid \theta^{i'})$ if $\theta^i > \theta^{i'}$.*

These are very similar to the standard Spence-Mirrlees condition (Mirrlees, 1971) in that they ensure that higher types always make decisions that result in higher incomes, and that welfare weights decline uniformly with income. However, they also imply that an individual with a higher type (θ_i) chooses a higher value of x_j for all j. In other words, individuals' decisions have the same ordering on every dimension. Lemma 2 formalizes this statement.

Lemma 2. *Given Assumption 4, $\boldsymbol{x}$ is monotone increasing in θ_i. Thus, an individual with higher θ_i has both higher x_j and higher $x_{j'}$ for all j and j'.*

This monotonicity implication is also an attribute of other analyses that feature multiple taxpayer decisions. For example Ferey et al. (2021) require it for their analysis of capital and labor income taxation.[29] Yet, we also acknowledge that it is restrictive. Nonetheless, without an assumption of this kind, general multi-dimensional heterogeneity makes the analysis of optimal taxation highly intractable.

The analysis is surprisingly tractable with these assumptions because it allows us to separately consider two distinct types of reform. First, we study redistribution from higher types to lower types. The result is an optimal tax condition that is similar to Mirrlees (1971) except that it features a correction for the internality. Second, we examine tax rates on different income-generating activities, holding fixed the overall amount of vertical redistribution. We find that they should be set to equalize a weighted average of the perceived an actual marginal tax rates across decisions, balancing the internality and externality in much the same way as we saw for optimal taxpayer education in Section 3.

4.1 The Optimal Amount of Redistribution

We start by studying redistribution from higher types to lower types. To do so, we consider a particular class of tax reform that has often been analyzed in the literature (see, e.g., Saez 2001): We raise the tax burden of individuals with income greater than some fixed threshold, y^*. This produces a small change in the marginal tax rate for individuals at y^* without any change to the marginal tax rate elsewhere. Specifically, individuals at y^* see their marginal tax rate raised by:

$$\tau_y\left(\boldsymbol{x}^*\right) d\eta \equiv \frac{\tau\left(\boldsymbol{x}^* + d\boldsymbol{x}\right) - \tau\left(\boldsymbol{x}^*\right)}{dy} d\eta = \frac{d\eta}{dy} \tag{26}$$

whereas all other individuals see no change in their marginal tax rates.

Because decision utility is quasilinear in consumption and the perceived tax rate at each level of income is unaffected by changes in the marginal tax rate elsewhere, only individuals at y^* respond. We can define the elasticity of income as:

$$\varepsilon_y^* = \left(\frac{1 - \mathrm{MTR}^*}{y^*}\right) \frac{dy_j^*}{-\tau_y\left(\boldsymbol{x}^*\right) d\eta} \tag{27}$$

where MTR^* is the average marginal tax rate at y^* (as defined in equation 5). The elasticity, ε_y^*, is a sufficient statistic for the welfare analysis below, but we note that its size and sign depend on how perceived marginal tax rates change when changes are made to the true tax system. We take this up further in Appendix A.2.

[29] Ferey et al. (2021) point out that an alternative way of viewing this is as a restriction on the class of tax systems, ruling out those which cause higher types to engage in more of some activities but less of others. In their context, for example, this is the set of tax systems under which higher earners save more.

The last piece of notation required is the cumulative distribution function for income, and the corresponding cumulative welfare function.

$$H(y) = \int_{\mathcal{I}:y^i \le y} di = \int_0^y h(v)dv \qquad \Psi(y) = \int_{\mathcal{I}:y^i \le y} \psi^i di = \int_0^y \overline{\psi}^v h(v)dv \tag{28}$$

where $h(y)$ is the density of income at y and $\overline{\psi}^y$ is the welfare weight of types with income y. Note that, because ψ^i is decreasing in θ^i and y^i is strictly increasing in θ^i, the cumulative welfare weight below any finite level of income is higher than the population weight. This in turn implies that $\Psi(y) \ge H(y)$, so that redistribution from high-income individuals to low-income individuals weakly raises welfare, all else equal.

Proposition 4 presents the main result regarding how optimal average marginal tax rates should be set at each level of income. Equation 29 is a first order condition for the planner that is similar to Mirrlees (1971), and is closely related to Gerritsen's (2016b) optimal tax condition when individuals choose only their overall level of income.

Proposition 4. *Given Assumptions 1, 3, 4 and 5, a necessary condition for the tax schedule to be optimal is that the average marginal tax rate at y^* is set such that equation 29 holds:*

$$\frac{(1-\psi^*)\text{MTR}^* + \psi^*\widetilde{\text{MTR}}^*}{1-\text{MTR}^*} = \frac{1}{\varepsilon_y^*}\frac{\Psi(y^*) - H(y^*)}{y^* h(y^*)} \tag{29}$$

where ψ^* is the welfare weight of individuals with income y^*; and MTR* and $\widetilde{MTR}^*$ are the average actual and perceived marginal tax rates at y^* (as defined in equation 5), respectively.

The main difference from Mirrlees (1971) is that the left-hand side of equation 29 contains an internality correction. If average actual and perceived marginal tax rates coincide, very little is new. But, in general, the planner balances a weighted average of the actual and perceived fiscal externalities in much the same way as for optimal taxpayer education in Section 3. Assuming that perceived marginal tax rates move in the same direction as actual tax rates when they are changed, a positive internality when taxpayers earn more (which again occurs when $\widetilde{\text{MTR}}^* > \text{MTR}^*$) means that marginal taxes are optimally set lower than the standard formula would imply, all else equal. This is because lower taxes stimulate effort, which has a benefit to taxpayers that they do not anticipate. Conversely, a negative internality (when $\widetilde{\text{MTR}}^* < \text{MTR}^*$) implies that higher taxes are optimal.

The extent to which the planner should take into account the internality depends on the welfare weight, ψ^*, of individuals at each level of income, y^*. If ψ^* is very low—which is often assumed to be the case at very high levels of income—the internality is heavily discounted, and only the fiscal externality matters. Equation 29 is largely isomorphic to

Mirrlees (1971) in that case. By contrast, if ψ^* is very high, the planner has a strong motive to keep actual and perceived marginal tax rates close to each other.

There are also more subtle differences from Mirrlees (1971) embodied in equation 29. First, the elasticity of income is a summary statistic for a multi-dimensional response across many activities. Second, the average marginal tax rates (true and perceived) weight together the responses of all the various decisions using the share of the response that comes from each decision. Both the elasticity of income and the average marginal tax rates differ depending on the specific perturbation being considered.

4.2 Relative Tax Rates

We next characterize the optimal tax treatment of different activities relative to each other, holding fixed the overall amount of redistribution from higher to lower types. Proposition 5 presents the main result in its most general form, which we build on below.

Proposition 5. *Consider an arbitrarily small reform that raises the tax burden on individuals with $x_j > x_j^*$ and reduces the tax burden by the same amount for those with $x_{j'} > x_{j'}^*$. Letting η index the size of the reform, the welfare impact is proportional to:*

$$\left[(1-\psi^*)\,MTR^* + \psi^*\widetilde{MTR}^*\right]\frac{dy^*}{d\eta} \tag{30}$$

Given Assumptions 1, 3, 4 and 5, a necessary condition for the tax schedule to be optimal is that expression 30 equals zero.

The policy change considered in Proposition 5 does not mechanically redistribute between between types. Nor does it change the rate at which the tax burden, $T(\boldsymbol{x})$ increases with total income, $y(\boldsymbol{x})$. However, it will in general lead individuals to re-optimize their decisions, because it changes the marginal tax rates that apply to decisions x_j and $x_{j'}$. This again produces a fiscal externality and an internality.

To understand the implications more deeply, it is helpful to start with a special case in which we adopt the same separability assumption as in Section 3.9 (Assumption 2) combined with an additional assumption (Assumption 6) that the marginal tax rate changes implied by Proposition 5 are equally salient.

Assumption 6. *The changes to the marginal tax rates on x_j and $x_{j'}$ in Proposition 5 are equally salient so that taxpayers do not (mistakenly) perceive that the reform mechanically changes the rate at which $T(\boldsymbol{x}^*)$ increases with $y(\boldsymbol{x}^*)$.*

These two assumptions jointly ensure that no type, θ^i, perceives any change to the amount of tax that they would pay if they adopted the same vector of decisions as before, or if they adopted the vector of decisions chosen by any other type. In turn, this means that the reform does not introduce an incentive for any type to change the level of income that they earn. Rather, the only thing it does is change how they earn that income.

In this special case in which no type changes their overall income, it can be shown that equation 30 holds when the weighted average, $W_j^* = (1-\psi^*)\,\mathrm{MTR}_j^* + \psi^* \widetilde{\mathrm{MTR}}_j^*$, is constant across the various decisions. To see why, we can simply rewrite equation 30 as the following summation:

$$\sum_{j \in \mathcal{J}} \left[(1-\psi^*)\,\mathrm{MTR}_j^* + \psi^* \widetilde{\mathrm{MTR}}_j^*\right] y_j\left(\boldsymbol{x}^*\right) \frac{dx_j^*}{d\eta} = \text{Constant} \times \frac{dy^*}{d\eta} = 0 \tag{31}$$

The first equality holds because $(1-\psi^*)\,\mathrm{MTR}_j^* + \psi^* \widetilde{\mathrm{MTR}}_j^*$ is a constant, while the second follows because incomes do not change. This result is formalized in Proposition 6 (i).

This is a very simple and important result. Its policy implications and their dependence on ψ^* echo our results for optimal taxpayer education. There are three key benchmarks. First, if taxpayers at y^* have a very low welfare weight—which may be true at very high incomes—the planner equalizes actual marginal tax rates across the income sources. This maximizes the revenue that can be extracted from these taxpayers. Second, *perceived* tax rates are equalized if the marginal social welfare weight is one, in which case the fiscal externality and internality are equally weighted. This ensures overall efficiency in the allocation of effort. Finally, if ψ^* is very high, the planner equalizes the internalities across the decisions, because reducing the harm from the internality is much more important than government revenue at the margin. Outside of these three situations, the weights are in between these three special cases.

Proposition 6. *Suppose $W_j^* = (1-\psi^*)\,MTR_j^* + \psi^* \widetilde{MTR}_j^*$ is equalized across decisions $j \in \mathcal{J}$.*

(i) If Assumption 2 (separability) and Assumption 6 (equal salience) both hold, then the welfare gain in expression 30 is zero.

(ii) If Assumption 2 holds but not Assumption 6, the welfare gain has the same sign as W_j^* if the change to the tax rate on x_j is less salient than the one for $x_{j'}$—and the opposite otherwise.

(iii) If Assumption 6 holds but not Assumption 2, the welfare gain has the same sign as W_j^* if higher types earn more of their marginal income via x_j rather than $x_{j'}$, compared to what lower types would choose if they increased their incomes—and the opposite otherwise.

Beside the special case in which Assumptions 2 and 6 hold, Proposition 6 provides two reasons, (ii) and (iii), to diverge from a tax system that equalizes the weighted average distortion in expression 30. First, if changes in the tax treatment of some decisions are more salient than others, individuals may *perceive* that there was an increase in the rate at which the tax burden increases with income even though there was not in reality. This would change the perceived incentive to earn a higher income. Assuming that the weighted distortion, W_j^*, is positive so that an increase in income raises welfare, the planner would therefore be able to reap an efficiency gain by shifting the tax burden toward decisions for which taxes are less salient: Doing so would preserve the amount of redistribution while causing individuals to perceive the tax system as less distortionary than before.

Second, individuals with different types may differ in their preferences regarding how to earn a given level of income. Incomes then change in response to the reform in Proposition 5, and it is again optimal to choose a tax system that does not equalize the distortion across activities. Assuming again that the weighted distortion (W_j^*) is positive, there is an efficiency gain from shifting the marginal tax burden toward activities that are intrinsically preferred by higher income individuals. Essentially, such innate differences in preferences allow the social planner to redistribute more efficiently by "tagging" high types using the composition of their incomes. It is precisely this type of tagging benefit that is ruled out by Assumption 2 (separability).

Example 4. *To consolidate these intuitions, consider a concrete example. Let $\boldsymbol{x}^i$ be comprised of labor supply (x_1^i), and working to upgrade a house to sell it for a capital gain (x_2^i). If Assumption 2 (separability) and Assumption 6 (equal salience) both hold, setting marginal tax rates such that equation 32 holds ensures that there is no welfare gain to changing them.*

$$(1-\psi^*)\, MTR_1^* + \psi^* \widetilde{MTR}_1^* = (1-\psi^*)\, MTR_2^* + \psi^* \widetilde{MTR}_2^* \tag{32}$$

For an individual with very low welfare weight (ψ^i close to zero), this implies that the true marginal tax rates on the two activities are equal, so that there is no tax advantage to shifting effort between x_1^i and x_2^i. For a taxpayer with the average social welfare weight ($\psi^i = 1$), perceived marginal tax rates are equalized, so that neither is perceived to be tax-favored. For those with very high welfare weights ($\psi^i \to \infty$), the gaps between the actual and perceived tax rates are equalized.

Relaxation of Assumption 6 changes this rule. For example, if an increase in the marginal tax rate on x_1^i is less salient than a decrease in the marginal tax rate on x_2^i, this introduces a reason to increase the marginal tax rate on x_1^i. Relaxation of Assumption 2 introduces an incentive to tax activities that are intrinsically preferred by higher income individuals.

4.3 Allowing for Non-local Responses to Tax Reforms

We now generalize the analysis to incorporate non-local responses to tax reforms. Unlike in simpler settings without misperceptions, there are two reasons why an individual may change her behavior in response to a reform that does not affect her marginal tax rates. First, there may be traditional income effects if we relax the quasilinearity of the utility function. Second, a taxpayer's perception of her marginal tax rates may change even if her true marginal tax rates do not.[30] Importantly, there is no need to distinguish between these two distinct mechanisms for the purpose of welfare analysis.

We allow each individual to respond to changes in the level of her tax burden as well as her marginal tax rate, replacing Assumption 3 with Assumption 7. This is sufficient to accommodate the "ironing" heuristic, in which an individual mistakenly responds to her average tax rate as if it were her marginal tax rate. This has no impact on our results regarding relative tax rates, but does affect the optimal amount of redistribution.[31]

Assumption 7. *The perceived marginal tax rate on decision j is a smooth function of k, $\boldsymbol{x}^i$, the true marginal tax rate on that decision, $T_j\left(\boldsymbol{x}^i, \eta\right)$, and an individual's tax burden, $T\left(\boldsymbol{x}^i, \eta\right)$.*

With this change to our assumptions, it becomes necessary to introduce two new pieces of notation. First, the compensated response of taxpayers at y^* to changes in their marginal tax rate (holding fixed their tax burden) is ε_c^*. Second, the "income effect" for taxpayer i (holding fixed her marginal tax rate) is η^i.

$$\varepsilon_c^* = \left(\frac{1-\text{MTR}^*}{y^*}\right) \left.\frac{dy_j^*}{-\tau_y\left(\boldsymbol{x}^*\right) d\eta}\right|_{\tau(\boldsymbol{x}^*)=0} \qquad \eta^i = \left(1-\text{MTR}^i\right) \left.\frac{dy^i}{-\tau(\boldsymbol{x}^i)d\gamma}\right|_{\tau_y(\boldsymbol{x}^i)=0} \tag{33}$$

Using this notation, the optimal amount of redistribution is given by Proposition 7.

Proposition 7. *With non-local responses, a necessary condition for $T(\boldsymbol{x}^i)$ to be optimal, given assumptions 1, 4, 5 and 7 is that the average marginal tax rate at y^* is such that equation 29 holds:*

$$\begin{aligned}\frac{(1-\psi^*)\text{MTR}^* + \psi^*\widetilde{\text{MTR}}^*}{1-\text{MTR}^*} &= \frac{1}{\varepsilon_c^*}\frac{\Psi\left(y^*\right) - H\left(y^*\right)}{y^* h\left(y^*\right)} \\ &\quad + \int_{y^*}^{\overline{y}} \frac{(1-\psi^y)MTR^y + \psi^y\widetilde{MTR}^y}{1-MTR^y}\frac{\eta^y h\left(y\right) dy}{\varepsilon_c^* h\left(y^*\right) y^*}\end{aligned} \tag{34}$$

where ψ^y is the welfare weight of individuals at y; MTR^y and $\widetilde{MTR}^y$ are the average actual and perceived marginal tax rates at y; and η^y is the income effect for taxpayers at y, respectively.

Equation 34 has very similar intuition to the version we obtained when we assumed that responses were local. However, it takes into account the fact that a higher marginal

[30] In practical applications, the size and composition of non-local responses to tax reforms will depend on the budget adjustment rule (see Reck 2016) that individuals follow when they misperceive the level of their tax burden, but this does not change the sufficient statistics formulas below.

[31] Note that the welfare weights must be interpreted more carefully with non-local responses: The average weight is not one, because a universal lump sum transfer induces a fiscal externality (and an internality).

tax rate at y^* increases the tax burden for individuals at higher levels of income. With traditional income effects, we would expect this to cause a positive fiscal externality as taxpayers at higher incomes increase their labor supply. However, this need not be the case here. For example, the ironing heuristic would suggest that incomes might fall. This is because a higher tax burden implies a higher average tax rate, which an ironing taxpayer mistakenly responds to as if it were her marginal tax rate.[32] An additional difference is that the impact of this re-optimization on social welfare again involves a weighted average of perceived and actual fiscal externalities.

5 Private Investment in Knowledge

Rather than the government directly choosing taxpayers' level of knowledge, one could imagine that individuals endogenously invest resources (cognitive and other) to discern the parameters of the tax system subject to government information provision policy.[33] We consider this possibility next.

5.1 Individual Knowledge Acquisition

Consider an individual who has beliefs about which of many tax systems, $T \in \boldsymbol{T}$, she might be facing. We assume that she faces a two-stage decision problem in response to a policy change. First, she chooses how much additional information to acquire. Then, she chooses $\boldsymbol{x}^i$, holding fixed her new level of knowledge.

The taxpayer starts with an initial belief, $\boldsymbol{p}_0$. Specifically, she subjectively assigns probability p_0^T to the tax system being T. Thus, her expected marginal tax rates are given by:

$$\widetilde{T}_j\left(\boldsymbol{x}^i \mid \boldsymbol{p}_0\right) = E_i\left(T_j\left(\boldsymbol{x}^i\right) \mid \boldsymbol{p}_0\right) = \sum_{T \in \boldsymbol{T}} p_1^T T_j\left(\boldsymbol{x}^i\right) \tag{35}$$

where $\boldsymbol{x}^i$ is the vector of income-generating activities she chooses.

The taxpayer makes an investment, a^i, to acquire more information in the form of a signal, $s^i \in [0,1]$. The signal has density $f\left(s^i \mid a^i, T\right)$, which we assume for simplicity is twice differentiable with respect to a^i. The cost of the investment, $\phi^i\left(a^i\right)/\beta$, is twice differentiable, strictly increasing and strictly convex. If the taxpayer invests nothing ($a^i = 0$), the signal comes at no cost but is not informative at all. The benefit of increasing a^i is

[32] See Rees-Jones and Taubinsky (2020) for a quantitative analysis of taxation with ironing taxpayers.

[33] Both mechanisms could operate simultaneously. Government information provision may then crowd out private knowledge acquisition, amounting to a transfer to those who acquire a substantial amount of it with little change in decisions. Even in this case, the same sufficient statistics apply.

that it increases the informativeness of s^i in the sense of Blackwell (1953). Upon receipt of the signal, the taxpayer revises her belief to $\boldsymbol{p}_1$, so that the probability she assigns to tax system T is p_1^T. Her decision vector changes from $\boldsymbol{x}_0^i$ to $\boldsymbol{x}_1^i(\boldsymbol{p}_1)$.

The parameter β is a tax system policy choice that lowers the individual cost of knowledge acquisition. For example, this could be a taxpayer education campaign, or an investment to make the IRS instructions and forms more accessible and understandable. The private cost of attaining a given amount of information, $\phi^i(a^i)/\beta$, may also vary across individuals of different types, depending on factors such as their cognitive ability.

The utility that a taxpayer of type θ^i will experience when uncertainty is resolved is:

$$y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i) - \frac{\phi^i(a^i)}{\beta} \tag{36}$$

Without acquiring any information (i.e., when $a^i = 0$), her expected utility is therefore:

$$\mathrm{EU}_0 = E_i\left[y(\boldsymbol{x}_0^i) - T(\boldsymbol{x}_0^i) - \chi(\boldsymbol{x}_0^i \mid \theta^i) \mid \boldsymbol{p}_0\right] \tag{37}$$

If she instead invests $a^i > 0$, her expected utility is:

$$\mathrm{EU}_1 = E_i\left\{E_i\left[y(\boldsymbol{x}_1^i(\boldsymbol{p}_1)) - T(\boldsymbol{x}_1^i(\boldsymbol{p}_1)) - \chi(\boldsymbol{x}_1^i(\boldsymbol{p}_1) \mid \theta^i) \mid \boldsymbol{p}_0\right]\right\} - \frac{\phi^i(a^i)}{\beta} \tag{38}$$

The return to investing $a^i > 0$ is $\mathrm{EU}_1 - \mathrm{EU}_0$.

We make two key assumptions to simplify the analysis from here. Assumption 8 states that there are diminishing marginal returns to acquiring more information about the tax system.[34] Assumption 9 amounts to assuming that taxpayers are initially optimizing before any policy change is made.

Assumption 8. *The cost function, ϕ^i, is convex enough such that $EU_1 - EU_0$ is strictly concave.*

Assumption 9. *In the absence of any policy change, the taxpayer is indifferent between remaining at $a^i = 0$ (acquiring no new information) or increasing a^i by an arbitrarily small amount.*

Jointly, these assumptions ensure that the net return to acquiring new information is strictly concave, and that it is zero in the absence of any policy change.

At the status quo when $a^i = 0$, we continue to assume that the taxpayer's optimal choice of $\boldsymbol{x}^i$ is pinned down by a first order condition.

$$y_j^i(\boldsymbol{x}_0^i) - \widetilde{T}_j(\boldsymbol{x}_0^i \mid \boldsymbol{p}_0) = \chi_j(\boldsymbol{x}_0 \mid \theta^i) \tag{39}$$

[34] It is straightforward to consider cases in which concavity is violated for $\boldsymbol{x}^i$ or a^i. This introduces the possibility that decisions jump discretely, but adds little conceptual insight.

The difference from Section 2 is that the individual's perceived marginal tax rates are explicitly *expected* marginal tax rates. In turn, because $f\left(s^i \mid a^i, T\right)$ is twice differentiable with respect to a^i, $\boldsymbol{x}^i$ is differentiable with respect to a^i.

Starting again from the status quo, differentiability allows us to write the net expected return to increasing a^i as:

$$\textbf{Net Expected Return} = \underbrace{E_i\Bigg(\sum_{j\in\mathcal{J}} \frac{dx_j^i}{da^i}\left[\widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) - T_j(\boldsymbol{x}^i)\right]\Bigg|\boldsymbol{p}_0\Bigg)}_{\textbf{Gross Expected Return}} - \underbrace{\frac{\phi_a^i(a^i)}{\beta}}_{\textbf{Cost}} \tag{40}$$

where dx_j^i/da^i is the total derivative of x_j^i with respect to a^i (evaluated at $a^i = 0$).

The benefit of additional knowledge is embodied by the gross return in equation 40. When a taxpayer invests, the law of iterated expectations implies that she cannot know the direction of the change in her decision vector, $\boldsymbol{x}^i$. It therefore follows that $E_i(dx_j^i/da^i) = 0$. However, she expects her perception of marginal tax rates to move closer to the truth, and her decisions to move closer to her true optimum. On average, then, we would expect that if $\widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) > T_j(\boldsymbol{x}^i)$, more information would cause $\widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0)$ to fall, and x_j to rise. Conversely, if $\widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) < T_j(\boldsymbol{x}^i)$, x_j would be expected to fall. For this reason, the gross expected return in equation 40 is greater than zero, which would be its value if dx_j^i/da^i and $\widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) - T_j(\boldsymbol{x}^i)$ were independent.

5.2 Knowledge Policy with Individual Investment

If the government knows that an individual's beliefs are biased in a particular direction, it can expect that bias to be reduced with knowledge acquisition. This implies a predictable change in perceptions from the government's point of view, which is unanticipated by the individual. This informational advantage forms the basis for optimal knowledge policy.

In the framework with endogenous information acquisition, government knowledge policy amounts to changing β, which changes the cost and therefore the attractiveness of individual knowledge acquisition. The social planner chooses this cost to maximize social welfare, which is defined similarly to before.

$$\max_{\beta\in\mathbb{R}^+} \boldsymbol{\mathcal{W}}(\beta, T) = \int_i \mathcal{W}\left(y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i) - \frac{\phi^i(a^i)}{\beta}\right) di \tag{41}$$

It does this subject to the constraint, also as before, that it must raise enough revenue to cover required government expenditure. This includes any direct cost of the information campaign, $C(\beta)$.

$$\int_i T(\boldsymbol{x}^i) di - C(\beta) \leq G \tag{42}$$

When the government makes knowledge acquisition cheaper by increasing β, there is again a fiscal externality and an internality, which can be aggregated in the same way as in Section 3. The main difference is that there is now also a mechanical cost saving for the individuals whose costs of knowledge investment have changed. This is offset by any mechanical cost to the government directly, $C(\beta)$. The total impact on social welfare is given by equation 43 of Proposition 8.

Proposition 8. *Consider an arbitrarily small perturbation that reduces the cost of knowledge acquisition by increasing β slightly. Given Assumptions 1, 8 and 9, the effect on social welfare is:*

$$\Delta\boldsymbol{\mathcal{W}} = \int_i \Bigg[\underbrace{(1-\psi^i)MTR_e^i E_g\left(\frac{dy^i}{d\beta}\right)}_{\substack{\textit{Actual Fiscal}\\ \textit{Externality}}} + \underbrace{\psi^i \widetilde{MTR}_e^i E_g\left(\frac{dy^i}{d\beta}\right)}_{\substack{\textit{Perceived Fiscal}\\ \textit{Externality}}}\Bigg] di + \underbrace{\int_i \psi^i \frac{\phi^i(a^i)}{\beta^2} di - C'(\beta)}_{\substack{\textit{Private and Government}\\ \textit{Direct Cost}}} \tag{43}$$

where MTR_e^i and $\widetilde{MTR}_e^i$ are the expected average marginal tax rates:

$$MTR_e^i = \frac{\sum_{j=1}^N MTR_j^i y_j E_g\left(\frac{dx_j^i}{d\beta}\right)}{\sum_{j=1}^N y_j E_g\left(\frac{dx_j^i}{d\beta}\right)} \qquad \widetilde{MTR}_e^i = \frac{\sum_{j=1}^N \widetilde{MTR}_j^i y_j E_g\left(\frac{dx_j^i}{d\beta}\right)}{\sum_{j=1}^N y_j E_g\left(\frac{dx_j^i}{d\beta}\right)} \tag{44}$$

and $E_g(dy^i/d\beta)$ and $E_g(dx_j^i/d\beta)$ are the government's expectations of individual behavioral responses when β is increased slightly.

Equation 43 is nearly identical to equation 13, and the models share many of the same implications. Specifically, the internality and fiscal externality feature additional expectation operators, but are qualitatively identical.[35] The other key difference is the introduction of a mechanical change to individual knowledge acquisition costs. These are weighted by the individual's marginal social welfare weights, which—assuming that welfare weights decline with income—makes it even more attractive than before to favor policies that increase the knowledge of poorer rather than richer taxpayers.

In any application, individual cost savings need to be traded off against direct costs to the government. For example, it may be costly to educate individuals about the tax credits for which they are eligible, but such a campaign would also save those individuals some of the time and effort they would otherwise have exerted to uncover that information. This mechanical benefit needs to be considered in addition to the internality and fiscal externality from any increase in knowledge that eventuates.

[35] Another small difference is that the derivatives here are directional. Behavior does not change if β is reduced, because the framework with knowledge acquisition does not allow knowledge to be revoked and taxpayer misled. One could adjust this if desired. For example, each of many generations of taxpayers could start from no knowledge rather than the status quo, and choose how much to acquire.

5.3 Optimal Taxation

While we have focused here on taxpayer education, the optimal tax problem would also be qualitatively unchanged if knowledge were endogenously acquired. Once a process is specified for how changes to the true tax system change taxpayers' beliefs, the equation governing the optimal amount of redistribution remains the same as above (equation 29), except that behavioral responses and marginal tax rates are again uncertain. The only further difference is that the response of income (e_z^*) derives from a somewhat different mechanism than in the model with direct government knowledge provision, which may have implications for its size and how it varies across individuals and reforms.

6 Conclusion

A large recent literature has addressed the implications of the fact that some taxpayers do not accurately perceive the tax system. This leads them to forego utility by making decisions based on a misperception of their true budget set, although this may be privately optimal if attention is costly to acquire. An entirely separate, and also large, literature has addressed the vast investments in knowledge, own or purchased, that people make to exploit the gray areas and fine points of tax systems in order to reduce their tax liability, legally and/or illegally.

This paper unifies these two central tax issues by recognizing that the tax system that some struggle to understand others strive to exploit. We develop a rigorous and general model that integrates the treatment of both phenomena, including the analysis of optimal policy. This model reveals that these two phenomena are isomorphic in important ways. The analysis clarifies the small set of aggregate sufficient statistics that are needed to quantify this approach, and how the parameters of an optimal tax system should reflect both the presence of incomplete knowledge of the true tax parameters as well as exploitation of the tax system. It is our view that this unified approach should replace the narrower conception of tax systems implicit in each of the two distinct literatures.

References

Abeler, Johannes and Simon Jäger. 2015. "Complex Tax Incentives." American Economic Journal: Economic Policy 7 (3): 1–28.

Akerlof, George A. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." American Economic Review 68 (1): 8–19.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2019. "Tax Evasion and Inequality." American Economic Review 109 (6): 2073–2103.

Bastani, Spencer, and Daniel Waldenström. 2021. "The Ability Gradient in Tax Tesponsiveness." Journal of Public Economics Plus 2: 1–19.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. "The Welfare Economics of Default Options in 401(k) Plans." American Economic Review 105 (9): 2798–2837.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." American Economic Review 105 (11): 3489–3529.

Battaglini, Marco, Luigi Guiso, Chiara Lacava, and Eleonora Patacchini. 2020. Tax Professionals and Tax Evasion. Working Paper.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." The Annals of Mathematical Statistics 24 (2): 265–272.

Blank, Joshua D., and Daniel Z. Levin. 2010. "When Is Tax Enforcement Publicized?" Virginia Tax Review 30 (1): 1–37.

Boccanfuso, Jérémy, and Antoine Ferey. 2021. "Inattention and the Taxation Bias". Working Paper.

Boning, William C., John Guyton, Ronald Hodge II, and Joel Slemrod. 2020. "Heard It Through the Grapevine: The Direct and Network Effects of a Tax Enforcement Field Experiment on Firms", Journal of Public Economics 190.

Bradford, David F. 1980. "The Economics of Tax Policy Towards Savings." In George M. Von Furstenberg (ed.), The Government and Capital Formation: 11–71. Cambridge, MA: Ballinger.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." American Economic Review 99 (4): 1145–1177.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," American Economic Review 103 (7).

Currie, Janet. 2006. "The Take-up of Social Benefits." In Alan J. Auerbach, David Card, and John M. Quigley (eds.), Public Policy and the Income Distribution: 80–148. New York: Russell Sage Foundation.

de Bartolme, Charles A. M. 1995. "Which Tax Rate Do People Use: Average or Marginal?", Journal of Public Economics 56: 79–96.

Fadlon, Itzik, and David Laibson. 2017. “Paternalism and Pseudo-Rationality.” National Bureau of Economic Research working paper #23260.

Farhi, Emmanuel, and Xavier Gabaix. 2020. "Optimal Taxation with Behavioral Agents." American Economic Review 110 (1): 298–336.

Ferey, Antoine, Benjamin B. Lockwood, and Dmitry Taubinsky. 2021. “Sufficient Statistics for Nonlinear Tax Systems with Preference Heterogeneity”, Working Paper.

Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." Review of Economics and Statistics 81 (4): 674–680.

Gerritsen, Aart. 2016a. “Optimal Nonlinear Taxation: The Dual Approach.” Max Planck Institute for Tax Law and Public Finance Working paper #2016–2.

Gerritsen, Aart. 2016b. “Optimal taxation when people do not maximize well-being.” Journal of Public Economics 144.

Goldin, Jacob. 2015. “Optimal Tax Salience.” Journal of Public Economics 131: 115–123.

Goldin, Jacob, and Daniel H. Reck. 2020. "Optimal Defaults with Normative Ambiguity." Review of Economics and Statistics: 1–45.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2021. "Tax Evasion at the Top of the Income Distribution: Theory and Evidence." National Bureau of Economic Research working paper #28542.

Hendren, Nathaniel. 2019. "Efficient Social Welfare Weights." Harvard University working paper.

Hoopes, Jeffrey L., Daniel H. Reck, and Joel Slemrod. 2015. "Taxpayer Search for Information: Implications for Rational Attention." American Economic Journal: Economic Policy 7 (3): 177-208.

Internal Revenue Service. 2019. Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013. Publication 1415. Washington, DC.

Internal Revenue Service. 2021. Internal Revenue Service Data Book, 2020. Publication 55-B. Washington, DC.

Keen, Michael, and Joel Slemrod. 2017. "Optimal Tax Administration." Journal of Public Economics 152: 133–142.

Kostøl, Andreas R. and Andreas S. Myhre. 2021. “Labor Supply Responses to Learning the Tax and Benefit Schedule.” American Economic Review 111 (11): 3733–66.

Liebman, Jeffrey B. and Erzo F. P. Luttmer. 2015. “Would People Behave Differently If

They Better Understood Social Security? Evidence from a Field Experiment." American Economic Journal: Economic Policy 7 (1): 275–99.

Liebman, Jeffrey, and Richard Zeckhauser. 2004. "Schmeduling." Harvard University working paper.

Lockwood, Benjamin B. 2020. "Optimal Income Taxation with Present Bias.", American Economic Journal: Economic Policy 12 (4): 298–327.

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl. 2017. "Taxation and the Allocation of Talent", Journal of Political Economy, 125(5): 1635-1682.

Moore, Dylan T., and Joel Slemrod. 2021. "Optimal Tax Systems with Endogenous Behavioral Biases." Journal of Public Economics, 197: 104384.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." Review of Economic Studies 38 (2): 175–208.

Morrison, William, and Dmitry Taubinsky. 2019. "Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes." National Bureau of Economic Research Working Paper #26180.

Moser, Christian, and Pedro Olea de Souza e Silva. 2019. "Optimal Paternalistic Savings Policies", SSRN Working Paper No. 17–51.

National Public Radio/Kaiser Family Foundation/Kennedy School of Government. 2003. National Survey of American' Views on Taxes. Available at https://www.kff.org/wp-content/uploads/2003/03/3340-t-survey-of-americans-views-on-taxes.pdf.

Persico, Nicola. 2000. "Information Acquisition in Auctions," Econometrica 68 (1), 135–148.

Reck, Daniel. 2016. "Taxes and Mistakes: What's in a Sufficient Statistic? ," Working Paper.

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring 'Schmeduling."' Review of Economic Studies 87 (5): 2399–2438.

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

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." Review of Economic Studies 85 (4): 2462–2496.

Topkis, Donald M. 1978. "Minimizing a Submodular Function on a Lattice." Operations Research 26 (2): 305–321.

A Supplementary Analysis
(For Online Publication)

A.1 Regularity for Optimal Knowledge

In this appendix, we discuss conditions such that decisions respond smoothly to the receipt of knowledge. We focus on sufficient conditions, but note that they are not necessary.

The first step is to ensure that a decision utility-maximizing choice exists, and the most straightforward way of achieving that is to assume that x_j is bounded below some maximum value, $\overline{x}_j$ for all $j \in \mathcal{J}$. Note that this assumption does not rule out corner solutions at which the first order condition for x_j does not hold. However, such corner solutions do not pose a problem for the analysis in Section 3.[36] The reason for this is that it does not matter for any of the welfare analysis whether the first order condition holds if x_j does not change in response to knowledge.

A more subtle requirement is that any interior decision utility-maximizing choice, $\boldsymbol{x}^*$, is unique. We first require that $\boldsymbol{x}^*$ *locally* maximizes utility, which in turn implies that $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is negative *semi-definite*. Throughout the paper, however, we strengthen this assuming that $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is negative *definite*. This allows us to apply the implicit function theorem, but we note that the assumption is not *required* for utility to be strictly concave at $\boldsymbol{x}^*$, and for $\boldsymbol{x}^*$ to respond smoothly to a perturbation. Beyond this, it remains possible that there are two distinct local optima that correspond to identical utility levels. In this case, a small change in knowledge could cause individuals to jump to a discretely different decision vector. We rule this out, but note that it would pose a problem for the analysis if it occurred for more than a zero measure set of the individuals under consideration. This clearly does not generically occur.

A.2 Anatomy of the Behavioral Response

An individual's behavioral response depends on the true tax system only via her perception of it. It is therefore important to understand how changes in the tax system are perceived. To study this, consider the same perturbation that we evaluated in Section 4. This was characterized by a size, η, and a direction, $\tau\left(\boldsymbol{x}^*\right)$. With this change to the tax schedule, the change in the marginal tax rate on x_j is given by equation 45.

$$\tau_j\left(\boldsymbol{x}^*\right) d\eta \equiv \frac{d\tau\left(\boldsymbol{x}^*\right)}{dx_j^*} d\eta \tag{45}$$

[36] It is straightforward to rule out corner solutions with assumptions on the tax schedule and utility.

The perceived tax system can be similarly parameterized. The size of the reform is still η, but the direction may differ. Specifically, we can write the perceived tax system as follows.

$$\widetilde{T}\left(\boldsymbol{x}^{*}, k, \eta\right)=\widetilde{\mathcal{T}}\left(\boldsymbol{x}^{*}, k\right)+\eta \tilde{\tau}\left(\boldsymbol{x}^{*}, k\right) \tag{46}$$

The change in the perceived marginal tax rate is then given by equation 47.

$$\widetilde{\tau}_{j}\left(\boldsymbol{x}^{*}\right) d\eta \equiv \frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{j}^{*}} d\eta \tag{47}$$

Responding to this change in perceived tax rates, individuals at y^{*} changes their vector of decisions as follows:

$$d\boldsymbol{x}^{*}=\mathcal{H}\left(\boldsymbol{x}^{*}\right)^{-1} \times d\widetilde{T}^{*}=\mathcal{H}\left(\boldsymbol{x}^{*}\right)^{-1}\left[\begin{array}{c}\frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{1}^{*}} d\eta \\ \vdots \\ \frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{|\mathcal{J}|}^{*}} d\eta\end{array}\right] \tag{48}$$

The change in income is therefore given by equation 49.

$$\left[\begin{array}{lll}y_{1}\left(\boldsymbol{x}^{*}\right) & \ldots & y_{|\mathcal{J}|}\left(\boldsymbol{x}^{*}\right)\end{array}\right] d\boldsymbol{x}^{*}=\left[\begin{array}{lll}y_{1}\left(\boldsymbol{x}^{*}\right) & \ldots & y_{|\mathcal{J}|}\left(\boldsymbol{x}^{*}\right)\end{array}\right] \mathcal{H}\left(\boldsymbol{x}^{*}\right)^{-1}\left[\begin{array}{c}\frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{1}^{*}} \\ \vdots \\ \frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{|\mathcal{J}|}^{*}}\end{array}\right] d\eta \tag{49}$$

The fiscal externality from this re-optimization is given by equation 50.

$$\left[\begin{array}{lll}T_{1}\left(\boldsymbol{x}^{i}\right) & \ldots & T_{|\mathcal{J}|}\left(\boldsymbol{x}^{*}\right)\end{array}\right] d\boldsymbol{x}^{*}=\left[\begin{array}{lll}T_{1}\left(\boldsymbol{x}^{*}\right) & \ldots & T_{|\mathcal{J}|}\left(\boldsymbol{x}^{*}\right)\end{array}\right] \mathcal{H}\left(\boldsymbol{x}^{*}\right)^{-1}\left[\begin{array}{c}\frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{1}^{*}} \\ \vdots \\ \frac{d\widetilde{\tau}\left(\boldsymbol{x}^{*}\right)}{dx_{|\mathcal{J}|}^{*}}\end{array}\right] d\eta \tag{50}$$

In general, neither the fiscal externality nor the change in income are easily signable. Even in the benchmark cases we consider below, in which perceived marginal tax rates change in particularly uniform ways, complementarity between different decisions can reverse the sign of behavioral impacts. This is an important difference from simpler settings in which individuals make a single decision such as how much to earn or consume. Beyond these cases, perceived tax rates need not follow actual tax rates in such a simple way. For example, introducing additional complexity into the system may make it more confusing so that perceived and actual rates drift further apart. Alternatively, individuals may pay more attention to higher tax rates, so that raising rates leads to an unexpectedly large negative fiscal externality. Or *any* change to the tax system might reduce k by making old knowledge obsolete. Nonetheless, two special cases are insightful.

Special Case 1. Suppose that we make a change to the true tax system such that perceived marginal tax rates fall in proportion to their existing marginal tax treatment, $T_j(\boldsymbol{x}^*)$. This would be the case, for example, if marginal tax rates on all decisions are equally salient, and the perturbation reduces $T_j(\boldsymbol{x}^*)$ proportionally across all activities. In this special case, the change in perceived marginal tax rates is:

$$\frac{d\widetilde{\tau}(\boldsymbol{x}^*)}{dx_j^*} = \lambda T_j(\boldsymbol{x}^*) \tag{51}$$

where λ is a constant.

This means that the impact of the behavioral response on tax revenue is as follows.

$$\mathbf{FE} = \begin{bmatrix} T_1(\boldsymbol{x}^*) & \ldots & T_{|\mathcal{J}|}(\boldsymbol{x}^i) \end{bmatrix} \mathcal{H}(\boldsymbol{x}^*)^{-1} \begin{bmatrix} T_1(\boldsymbol{x}^*) \\ \vdots \\ T_{|\mathcal{J}|}(\boldsymbol{x}^*) \end{bmatrix} \lambda d\eta \tag{52}$$

Since $\mathcal{H}(\boldsymbol{x}^*)^{-1}$ is negative definite, this has the opposite sign of sign of λ. Thus, if perceived rates increased from a positive level, government revenue falls from this behavioral response. In summary, a perceived proportional reduction in marginal tax rates leads to a predictably positive fiscal externality.

Even within this special case, misperception of marginal tax rates may have important implications. For example, individuals may under-react or over-react to changes in marginal tax rates. If they under-react, then λ may be very small. In turn, this implies a small fiscal externality. Over-reaction is also a possibility.

Special Case 2. If the true tax system is changed such that perceived marginal tax rates fall in proportion to the slope of the income function, $y_j(\boldsymbol{x}^*)$, then the income response is similarly signable. Finally, note that for a harmonized tax system in which the marginal tax rate, $T_j(\boldsymbol{x}^*)/y_j(\boldsymbol{x}^*)$ is constant across all activities, the two special cases are identical. The fiscal externality and income response are then signable under the same circumstances: Intuitively, the sign of the income response suffices to calculate the fiscal externality if all activities are taxed at the same rate.

B Proofs

(For Online Publication)

Proof of Proposition 1. The unidimensional case is most straightforward. The first order impact of knowledge on individual welfare in this case is:

$$\begin{aligned}\mathbf{IN} &= \psi^i \left[\widetilde{\mathrm{MTR}}_1^i - \mathrm{MTR}_1^i\right] y_1(\boldsymbol{x}^i) dx_1^i \\ &= \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times \psi^i \left[\widetilde{T}_1(\boldsymbol{x}^i \mid k, T) - T_1(\boldsymbol{x}^i)\right] \times \frac{\partial \widetilde{T}_1}{\partial k} \qquad (53)\end{aligned}$$

Since $\boldsymbol{x}^i$ has only one element, $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is simply the second order condition. It is strictly negative at a local maximum. The welfare weight is positive. Finally, $\widetilde{T}_1(\boldsymbol{x}^i \mid k, T) - T_1(\boldsymbol{x}^i)$ is assumed to have the opposite sign of $\frac{\partial \widetilde{T}_1}{\partial k}$. Thus, **IN** is positive in equation 53.

Next consider the case in which utility is additively separable in each decision. In this case, $\mathcal{H}^i\left(\boldsymbol{x}^i\right)$ is diagonal, and so is its inverse. In turn, we are left with an equation equivalent to 53 for every decision, j.

$$\mathbf{IN} = \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times \psi^i \left[\widetilde{T}_j(\boldsymbol{x}^i \mid k, T) - T_j(\boldsymbol{x}^i)\right] \times \frac{\partial \widetilde{T}_j}{\partial k} \qquad (54)$$

Precisely the same logic as before applies here, so that **IN** is positive.

Finally, consider the case in which $\widetilde{\mathrm{MTR}}_j^i - \mathrm{MTR}_j^i$ is independent of j for any value of k, and define $M(k) = \widetilde{\mathrm{MTR}}_j^i - \mathrm{MTR}_j^i$ as that wedge.

$$\begin{aligned}\mathbf{IN} \times dk &= \psi^i \sum_{j \in \mathcal{J}} \left[\widetilde{\mathrm{MTR}}_j^i - \mathrm{MTR}_j^i\right] y_j(\boldsymbol{x}^i) dx_j^i \qquad (55) \\ &= \psi^i M(k) \times \sum_{j \in \mathcal{J}} y_j(\boldsymbol{x}^i) dx_j^i \qquad (56)\end{aligned}$$

Next, observe that $M(k)$ being constant across all decisions for all k implies that $\frac{\partial \widetilde{T}_j}{\partial k} / y_j$ is constant across all decisions. Define C as that constant. Letting $\widetilde{\boldsymbol{y}} = \left[y_1(\boldsymbol{x}^i), \ldots, y_{|\mathcal{J}|}(\boldsymbol{x}^i)\right]$, we can write $\mathbf{IN} \times dk$ as:

$$\mathbf{IN} \times dk = \psi^i M(k) C \times \widetilde{\boldsymbol{y}} \times \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times \widetilde{\boldsymbol{y}}' \qquad (57)$$

Since $\boldsymbol{x}^i$ is a locally unique utility maximizing choice, $\mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1}$ is negative definite. This implies that $\widetilde{\boldsymbol{y}} \times \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times \widetilde{\boldsymbol{y}}'$ is weakly negative. Just as before, the welfare weight is positive, while $M(k)$ and C have opposite signs. Thus, **IN** is weakly positive. $\square$

Proof of Proposition 2. We start from social welfare as defined in equation 6 of Section 2, which we restate here for convenience:

$$\boldsymbol{\mathcal{W}}(k,T) = \int_i \mathcal{W}\left(y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i)\right) di - C(k) \tag{58}$$

Assumption 1 ensures that $\boldsymbol{x}^i$ responds smoothly as k changes. Thus, we can differentiate $\boldsymbol{\mathcal{W}}(k,T)$ with respect to k to obtain the direct first-order impact of a marginal increase in knowledge on social welfare. When we do this, we obtain:

$$\int_i \mathcal{W}'(U^i) \times \sum_{j\in\mathcal{J}} \left(y^i_j(\boldsymbol{x}^i) - \widetilde{T}_j(\boldsymbol{x}^i \mid k,T) - \chi_j(\boldsymbol{x}^i \mid \theta^i)\right) \frac{dx^i_j}{dk} di \tag{59}$$

Combining this with the first order condition (equation 3) and using the definition of the welfare weight, we obtain the internality.

$$\lambda \int_i \psi_i \sum_{j\in\mathcal{J}} \left(T_j(\boldsymbol{x}^i) - \widetilde{T}_j(\boldsymbol{x}^i \mid k,T)\right) \frac{dx^i_j}{dk} di \tag{60}$$

Converting this to government dollars by dividing by λ, and using the definitions of the actual and perceived average marginal tax rates, we obtain the internality term (integrated over the population).

$$\int_i \psi_i \left(\mathrm{MTR}^i(\boldsymbol{x}^i) - \widetilde{\mathrm{MTR}}^i(\boldsymbol{x}^i \mid k,T)\right) \frac{dy^i}{dk} di \tag{61}$$

Next, there is the impact on government revenue via the resource constraint, which is comprised of the direct cost and the fiscal externality.

$$\int_i \left(\sum_{j\in\mathcal{J}} T_j(\boldsymbol{x}^i \mid k,T) \frac{dx^i_j}{dk}\right) di - C'(k) \tag{62}$$

Using the definition of the average actual marginal tax rate, we obtain:

$$\int_i \mathrm{MTR}^i(\boldsymbol{x}^i) \frac{dy^i}{dk} di - C'(k) \tag{63}$$

Combining this with the internality term, we obtain the change in welfare, $\Delta\boldsymbol{\mathcal{W}}$, stated in Proposition 2 (equation 13). □

Proof of Example 2. The only statement in example 2 that requires proof is that $dx_2/dk > 0$, since equation 18 follows directly from equation 8.

In this example, totally differentiating the first order conditions leads to the following equation for the change in decisions.

$$\begin{bmatrix} y_{11} - \widetilde{T}_{11} - \chi_{11} & y_{12} - \widetilde{T}_{12} - \chi_{12} \\ y_{12} - \widetilde{T}_{12} - \chi_{12} & y_{22} - \widetilde{T}_{22} - \chi_{22} \end{bmatrix} \begin{bmatrix} \frac{dx_1}{dk} \\ \frac{dx_2}{dk} \end{bmatrix} = \begin{bmatrix} \frac{\partial \widetilde{T}_1}{\partial k} \\ \frac{\partial \widetilde{T}_2}{\partial k} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{\partial \widetilde{T}_2}{dk} \end{bmatrix} \tag{64}$$

Inverting the matrix on the left, we obtain equations for the changes in x_1 and x_2.

$$\frac{dx_1}{dk} = \frac{-y_{12} + \widetilde{T}_{12} + \chi_{12}}{\left(y_{11} - \widetilde{T}_{11} - \chi_{11}\right)\left(y_{22} - \widetilde{T}_{22} - \chi_{22}\right) - \left(y_{12} - \widetilde{T}_{12} - \chi_{12}\right)^2} \left(\frac{\partial \widetilde{T}_2}{\partial k}\right) \tag{65}$$

$$\frac{dx_2}{dk} = \frac{y_{11} - \widetilde{T}_{11} - \chi_{11}}{\left(y_{11} - \widetilde{T}_{11} - \chi_{11}\right)\left(y_{22} - \widetilde{T}_{22} - \chi_{22}\right) - \left(y_{12} - \widetilde{T}_{12} - \chi_{12}\right)^2} \left(\frac{\partial \widetilde{T}_2}{\partial k}\right) \tag{66}$$

Since $\mathcal{H}\left(\boldsymbol{x}^i\right)$ is negative definite, the denominator in both expressions is positive, while the numerator of $dx_2/dk > 0$ is negative. We know that $\partial \widetilde{T}_2/\partial k < 0$ because $T_2(\boldsymbol{x}^i) < \widetilde{T}_2(\boldsymbol{x}^i \mid k, T)$. Thus, $dx_2/dk > 0$. $\square$

Proof of Example 3. First, observe that we can rewrite the wedge, $\omega_j^i = \widetilde{\text{MTR}}_j^i - \text{MTR}_j^i$ as a function solely of $\widetilde{\text{MTR}}_j^i$ and $\overline{\text{MTR}}^i$.

$$\omega_j^i = \widetilde{\text{MTR}}_j^i - \text{MTR}_j^i = (1-d)\,\widetilde{\text{MTR}}_j^i - (1-d)\,\overline{\text{MTR}}^i \tag{67}$$

Next, recall that the change in the decision vector, $\boldsymbol{x}^i$, is:

$$d\boldsymbol{x}^i = \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times d\widetilde{T}^i \tag{68}$$

Since knowledge proportionally reduces ω_j^i, the jth element of $d\widetilde{T}^i = -y_j(\boldsymbol{x}^i) \times \omega_j^i \times r$ for some constant r, $0 < r < 1$. Since $\psi^i = 1$, the effect on welfare is:

$$-\left[\widetilde{\text{MTR}}_1^i y_1(\boldsymbol{x}^i), \ldots, \widetilde{\text{MTR}}_{|\mathcal{J}|}^i y_{|\mathcal{J}|}(\boldsymbol{x}^i)\right] \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \left[y_1(\boldsymbol{x}^i)\omega_1^i, \ldots, y_{|\mathcal{J}|}(\boldsymbol{x}^i)\omega_{|\mathcal{J}|}^i\right]' (1-d)r \tag{69}$$

Next, there is a fiscal externality from any change in i's overall income:

$$\overline{\text{MTR}}^i \times \left[y_1(\boldsymbol{x}^i), \ldots, y_{|\mathcal{J}|}(\boldsymbol{x}^i)\right] \times \mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1} \times d\widetilde{T}^i = \overline{\text{MTR}}^i \times dy \tag{70}$$

Subtracting this from the change in welfare in equation 69 and adding it back in yields:

$$\Delta\boldsymbol{\mathcal{W}} = \underbrace{-\widetilde{\boldsymbol{y}}\mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1}\widetilde{\boldsymbol{y}}'(1-d)r}_{\textbf{Reallocation}} + \underbrace{\overline{\text{MTR}}^i \times dy}_{\textbf{Overall}} \tag{71}$$

where:

$$\widetilde{\boldsymbol{y}} = \left[y_1(\boldsymbol{x}^i)(\widetilde{\text{MTR}}_1^i - \overline{\text{MTR}}^i), \ldots, y_{|\mathcal{J}|}(\boldsymbol{x}^i)(\widetilde{\text{MTR}}_{|\mathcal{J}|}^i - \overline{\text{MTR}}^i)\right] \tag{72}$$

$\mathcal{H}^i\left(\boldsymbol{x}^i\right)^{-1}$ is negative definite because $\boldsymbol{x}^i$ is the locally utility maximizing choice; and $0 < r < 1$ by assumption. Thus, the reallocation term in equation 71 is positive if $d < 1$, negative if $d > 1$, and zero if $d = 1$. The fiscal externality from overall effort is proportional to the change in overall income, and zero if $dy = 0$. This concludes the proof. □

Proof of Lemma 1. Maximizing over possible choices of x^i amounts to the following Lagrangean.

$$\max_{\boldsymbol{x}^i} y\left(\boldsymbol{x}^i\right) - \widetilde{T}\left(\boldsymbol{x}^i\right) - \chi\left(\boldsymbol{x}^i\right) - \widetilde{\chi}\left(y\left(\boldsymbol{x}^i\right) \mid \theta^i\right) - \lambda y\left(\boldsymbol{x}^i\right)$$

The first order condition for choice j is $y_j - \widetilde{T}_j - \chi_j - \chi_y^i y_j = \lambda y_j$. Re-arranging, this implies that the trade-off between good j and good j' is pinned down by: $(\widetilde{T}_1 + \chi_1)/(\widetilde{T}_2 + \chi_2) = y_1/y_2$. This is independent of i. □

Proof of Proposition 3. The proof will involve three steps. First, the perceived tax schedule will be replaced with another which is harmonized, and which leads to the same perceived utility at each level of income for all individuals. Second, the true tax schedule will be adjusted so that true utility is the same for each individual as it was before harmonization. Third, we show that revenue strictly increases for the government. This additional revenue can then be returned as a lump sum to construct a strict Pareto gain.

<u>Step 1:</u> Let $V_i(\widetilde{T})$ be the perceived indirect utility of individual i given the original perceived tax schedule $\widetilde{T}(\boldsymbol{x}^i)$:

$$V_i(\widetilde{T}) = \max_{\boldsymbol{x}} y(\boldsymbol{x}) - \widetilde{T}(\boldsymbol{x}) - \chi\left(\boldsymbol{x}\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}\right) \mid \theta^i\right) \tag{73}$$

Let x^* be the solution to that, so that income is $y\left(\boldsymbol{x}^*\right)$ and the tax burden is $\widetilde{T}(\boldsymbol{x}^*)$.

Starting from that point, replace $\widetilde{T}(\boldsymbol{x})$ with a harmonized perceived tax schedule $\widetilde{T}^\circ(y(\boldsymbol{x}))$ and let $V_i(\widetilde{T}^\circ)$ be the perceived indirect utility given this new tax system.

$$V_i(\widetilde{T}^\circ) = \max_{\boldsymbol{x}} y\left(\boldsymbol{x}\right) - \widetilde{T}^\circ\left(y\left(\boldsymbol{x}\right)\right) - \chi\left(\boldsymbol{x}\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}\right) \mid \theta^i\right) \tag{74}$$

As before, let $\boldsymbol{x}^\dagger$ be the solution. Income is $y(\boldsymbol{x}^\dagger)$. The perceived tax burden is $\widetilde{T}^\circ(y(\boldsymbol{x}^\dagger))$.

Observe that there are many such harmonized perceived tax schedules. We choose the unique such tax schedule such that $V_i(\widetilde{T}^\circ) = V_i(\widetilde{T})$ at the same level of income, $y(\boldsymbol{x}^*)$. Clearly, the focal individual is then indifferent and has no incentive to change their income level. Moreover, since any other individual would choose the same vector of decisions $\boldsymbol{x}$ if they were to earn $y(\boldsymbol{x}^*)$, this replacement of the tax schedule does not provide any individual with an incentive to change their income level.

Step 2: The new perceived marginal tax rate is the same across the various decisions under this new tax system (unlike before). This will ensure that the individual earns the same income more efficiently than before. Our next step will be to show that this in turn leads *perceived* government revenue to increase. The higher level of perceived government revenue can then be used to construct a perceived Pareto gain.

To prove this, suppose (subject to contradiction) that some individual can get at least as much perceived utility given the same choices as before, $\boldsymbol{x}^*$. Then they would get:

$$y\left(\boldsymbol{x}^*\right) - \widetilde{T}^\circ\left(y\left(\boldsymbol{x}^*\right)\right) - \chi\left(\boldsymbol{x}^*\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^*\right) \mid \theta^i\right) \tag{75}$$

But their first order conditions are different given this new tax schedule. They would pick a different bundle (i.e., $\boldsymbol{x}^\dagger \neq \boldsymbol{x}^*$) and their utility would be strictly higher. That is a contradiction, because we constructed this tax schedule such that $V_i(\widetilde{T}^\circ) = V_i(\widetilde{T})$. Thus, they must perceive that they would have lower utility given their old decision vector.

$$y\left(\boldsymbol{x}^*\right) - \widetilde{T}^\circ\left(y\left(\boldsymbol{x}^*\right)\right) - \chi\left(\boldsymbol{x}^*\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^*\right) \mid \theta^i\right) < V_i(\widetilde{T}) \tag{76}$$

Since $V_i(\widetilde{T}) = V_i(\widetilde{T}^\circ) = \max_{\boldsymbol{x}} y(\boldsymbol{x}^*) - \widetilde{T}^\circ(y(\boldsymbol{x}^*)) - \chi\left(\boldsymbol{x}^*\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^*\right) \mid \theta^i\right)$ we have

$$y(\boldsymbol{x}^*) - \widetilde{T}^\circ(y(\boldsymbol{x}^*)) - \chi\left(\boldsymbol{x}^*\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^*\right) \mid \theta^i\right) < y(\boldsymbol{x}^*) - \widetilde{T}(\boldsymbol{x}^*) - \chi\left(\boldsymbol{x}^*\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^*\right) \mid \theta^i\right) \tag{77}$$

which implies that $\widetilde{T}(\boldsymbol{x}^*) < \widetilde{T}^\circ(y(\boldsymbol{x}^*))$. Thus, since $y(\boldsymbol{x}^*) = y(\boldsymbol{x}^\dagger)$, we have our result that perceived tax revenue is higher under the neutral tax system: $\widetilde{T}^\circ(y(\boldsymbol{x}^\dagger)) > \widetilde{T}(\boldsymbol{x}^*)$. This is true for all individuals, so it is true in aggregate. Finally we use the perceived surplus revenue to provide a non-distortionary lump sum transfer, $\overline{T}$, that increases utility for all individuals. Letting $\widetilde{T}^\dagger(y(\boldsymbol{x}^\dagger)) = \widetilde{T}^\circ(y(\boldsymbol{x}^\dagger)) - \overline{T}$, we have constructed a *perceived* Pareto gain without a perceived decrease in government revenue.

Step 3: Our final step is to construct a true Pareto gain in addition to the perceived again. The process is very similar to Step 2. First observe that the perceived utility of each individual is equal under $\widetilde{T}^{\circ}(y(\boldsymbol{x}^{\dagger}))$ and $\widetilde{T}(\boldsymbol{x}^{*})$. This implies that:

$$\begin{aligned} y(\boldsymbol{x}^{\dagger}) - \widetilde{T}^{\circ}(y(\boldsymbol{x}^{\dagger})) - \chi(\boldsymbol{x}^{\dagger}) - \widehat{\chi}(y(\boldsymbol{x}^{\dagger}) \mid \theta^{i}) =& y(\boldsymbol{x}^{*}) - \widetilde{T}(\boldsymbol{x}^{*}) - \chi\left(\boldsymbol{x}^{*}\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^{*}\right) \mid \theta^{i}\right) \\ \widetilde{T}^{\circ}(y(\boldsymbol{x}^{\dagger})) - \widetilde{T}(\boldsymbol{x}^{*}) =& y(\boldsymbol{x}^{\dagger}) - \chi(\boldsymbol{x}^{\dagger}) - \widehat{\chi}(y(\boldsymbol{x}^{\dagger}) \mid \theta^{i}) \\ & - y(\boldsymbol{x}^{*}) + \chi\left(\boldsymbol{x}^{*}\right) + \widehat{\chi}\left(y\left(\boldsymbol{x}^{*}\right) \mid \theta^{i}\right) \end{aligned} \tag{78}$$

Since $\widetilde{T}^{\circ}(y(\boldsymbol{x}^{\dagger})) > \widetilde{T}(\boldsymbol{x}^{*})$, we know that:

$$y(\boldsymbol{x}^{\dagger}) - \chi(\boldsymbol{x}^{\dagger}) - \widehat{\chi}(y\left(\boldsymbol{x}^{\dagger}) \mid \theta^{i}\right) - y(\boldsymbol{x}^{*}) + \chi\left(\boldsymbol{x}^{*}\right) + \widehat{\chi}\left(y\left(\boldsymbol{x}^{*}\right) \mid \theta^{i}\right) > 0 \tag{79}$$

Now, the requirement for the individual's true utility to be unchanged is:

$$\begin{aligned} y(\boldsymbol{x}^{\dagger}) - T^{\dagger}(y(\boldsymbol{x}^{\dagger})) - \chi(\boldsymbol{x}^{\dagger}) - \widehat{\chi}(y(\boldsymbol{x}^{\dagger}) \mid \theta^{i}) =& y(\boldsymbol{x}^{*}) - T(\boldsymbol{x}^{*}) - \chi\left(\boldsymbol{x}^{*}\right) - \widehat{\chi}\left(y\left(\boldsymbol{x}^{*}\right) \mid \theta^{i}\right) \\ T^{\dagger}(y(\boldsymbol{x}^{\dagger})) - T(\boldsymbol{x}^{*}) =& y(\boldsymbol{x}^{\dagger}) - \chi(\boldsymbol{x}^{\dagger}) - \widehat{\chi}(y(\boldsymbol{x}^{\dagger}) \mid \theta^{i}) \\ & - y(\boldsymbol{x}^{*}) + \chi\left(\boldsymbol{x}^{*}\right) + \widehat{\chi}\left(y\left(\boldsymbol{x}^{*}\right) \mid \theta^{i}\right) > 0 \end{aligned}$$

where the final inequality follows from above. This implies that $T^{\dagger}(y(\boldsymbol{x}^{\dagger})) > T(\boldsymbol{x}^{*})$. In other words, the government can raise additional revenue from each individual and leave them equally well off. This is true for all individuals, and thus also in the aggregate. Finally, it can use the surplus revenue to construct a real Pareto gain. □

Proof of Lemma 2. This is a straightforward application of Topkis (1978). □

Proof of Proposition 4. We start by characterizing the effects of an arbitrary tax reform of size γ. Such a reform may change each individual's tax burden, their marginal tax rate and government revenue. We first compute these impacts, and then study the implications for welfare.

<u>Tax burden:</u> First, the effect of the reform on individual i's tax burden is:

$$\frac{dT\left(\boldsymbol{x}^{i},\gamma\right)}{d\gamma} = \tau\left(\boldsymbol{x}^{i}\right) + \sum_{j\in J} T_{j}\left(\boldsymbol{x}^{i},\gamma\right)\frac{dx_{j}^{i}}{d\gamma} \tag{80}$$

where $T_{j}\left(\boldsymbol{x}^{i},\gamma\right)$ is individual i's initial marginal tax rate with respect to decision x_{j}.

Marginal tax rates: The effect on this individual's marginal tax rate for decision j is:

$$\frac{dT_j\left(\boldsymbol{x}^i,\gamma\right)}{d\gamma} = \tau_j\left(\boldsymbol{x}^i\right) + \sum_{j\in J}\sum_{j'\in J} T_{jj'}\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma} \tag{81}$$

where $T_{jj'}\left(\boldsymbol{x}^i,\gamma\right)$ is the effect on this individual's marginal tax rate for decision j when decision j' increases.

As above, these summations take into account the fact that the effect of the reform on tax liability and the marginal tax rate depend on the responses of all decisions, in addition to the direct mechanical effect of the reform itself.

Government revenue: The government budget constraint is:

$$\mathcal{B} = \int_{\mathcal{I}} T\left(\boldsymbol{x}^i,\gamma\right) di \tag{82}$$

so the effect on government revenue is as follows:

$$\frac{d\mathcal{B}}{d\gamma} = \int_{\mathcal{I}}\left[\tau\left(\boldsymbol{x}^i\right) + \sum_{j\in J} T_j\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma}\right] di \tag{83}$$

Welfare: The direct impact on welfare of the reform is:

$$-\int_{\mathcal{I}}\psi^i\left(\tau\left(\boldsymbol{x}^i\right) + \sum_{j\in J}\left(T_j\left(\boldsymbol{x}^i,\gamma\right) - \widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)\right)\frac{dx_j^i}{d\gamma}\right) \tag{84}$$

and the total effect on welfare, including the fiscal externality, is:

$$\begin{aligned}
\frac{d\mathcal{W}}{d\gamma} &= \int_{\mathcal{I}}\left[\tau\left(\boldsymbol{x}\right) + \sum_j T_j\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma}\right] di \\
&\quad - \int_{\mathcal{I}}\psi^i\left(\tau\left(\boldsymbol{x}^i\right) + \sum_{j\in J}\left(T_j\left(\boldsymbol{x}^i,\gamma\right) - \widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)\right)\frac{dx_j^i}{d\gamma}\right) di \\
&= \int_{\mathcal{I}}\left(1-\psi^i\right)\tau\left(\boldsymbol{x}^i\right) \\
&\quad + \underbrace{\int_{\mathcal{I}}\left[\sum_j T_j\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma}\right] di}_{\textbf{Fiscal Externality}} - \underbrace{\int_{\mathcal{I}}\psi^i\left(\sum_{j\in J}\left(T_j\left(\boldsymbol{x}^i,\gamma\right) - \widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)\right)\frac{dx_j^i}{d\gamma}\right) di}_{\textbf{Internality}}
\end{aligned} \tag{85}$$

Specific perturbation: We next study the optimal amount of redistribution by studying a particular perturbation which raises revenue from the rich and returns that revenue as a lump sum (after netting out any loss due to the fiscal externality). Specifically, let $\tau\left(\boldsymbol{x}\right)=1$ if $\boldsymbol{x}>\boldsymbol{x}^{*}$ and zero otherwise. This produces a small change in the marginal tax rate for individuals at y^{*} without any change to the marginal tax rate elsewhere. Individuals at y^{*} see their tax rate raised by:

$$\tau_{y}\left(\boldsymbol{x}^{*}\right)d\gamma\equiv\frac{\tau\left(\boldsymbol{x}^{*}+d\boldsymbol{x}\right)-\tau\left(\boldsymbol{x}^{*}\right)}{dy}d\gamma=\frac{d\gamma}{dy} \tag{86}$$

whereas all other individuals see no change in their marginal tax rates.

Since decision utility is quasilinear in consumption and the perceived tax rate at each level of income is unaffected by changes in the marginal tax rate elsewhere, only individuals at y^{*} respond. By contrast, resources are taken from all individuals with greater than that level of income. Thus, we can rewrite the change in welfare as:

$$\frac{d\mathcal{W}}{d\gamma}=\int_{\mathcal{I}:y^{i}>y^{*}}\left(1-\psi^{i}\right)di+\int_{\mathcal{I}:y^{i}=y^{*}}\left[\sum_{j}T_{j}\left(\boldsymbol{x}^{i},\gamma\right)\frac{dx_{j}^{i}}{d\gamma}\right]di \tag{87}$$

$$-\int_{\mathcal{I}:y^{i}=y^{*}}\psi^{i}\left(\sum_{j\in J}\left(T_{j}\left(\boldsymbol{x}^{i},\gamma\right)-\widetilde{T}\left(\boldsymbol{x}^{i},k,\gamma\right)\right)\frac{dx_{j}^{i}}{d\gamma}\right)di \tag{88}$$

At any given level of income, we know that individuals are identical because $\boldsymbol{x}^{i}$ increases strictly with θ^{i}. Combining this fact with the definitions of MTR^{i} and $\widetilde{\mathrm{MTR}}^{i}$, we can write the actual and perceived marginal tax rates of individuals at y^{*} as MTR^{*} and $\widetilde{\mathrm{MTR}}^{*}$. Similarly, let the change in income at y^{*} be $dy^{*}/d\gamma$ and the welfare weight at that level of income be ψ^{*}. Then we can rewrite the welfare change as:

$$\frac{d\mathcal{W}}{d\gamma}=\int_{\mathcal{I}:y^{i}>y^{*}}\left(1-\psi^{i}\right)di+\int_{\mathcal{I}:y^{i}=y^{*}}\left[(1-\psi^{*})\mathrm{MTR}^{*}+\psi^{*}\widetilde{\mathrm{MTR}}^{*}\right]\frac{dy^{*}}{d\gamma}di \tag{89}$$

We obtain the final expression for the change in welfare by rewriting this in terms of the income distribution and using the definitions of ε_{y}^{*}, $H\left(y\right)$, $\Psi\left(y\right)$ and $h(y)$.

$$\frac{d\mathcal{W}}{d\gamma}=\Psi\left(y^{*}\right)-H\left(y^{*}\right)-\frac{(1-\psi^{*})\mathrm{MTR}^{*}+\psi^{*}\widetilde{\mathrm{MTR}}^{*}}{1-\mathrm{MTR}^{*}}\times\varepsilon_{y}^{*}y^{*}h\left(y^{*}\right) \tag{90}$$

Setting this welfare change equal to zero yields the equation in the proposition. □

Proof of Proposition 5. How should the marginal tax rate on various decisions compare to each other, holding the distribution of resources across individuals fixed?

To address this, consider a level of income, y^*, which corresponds to a unique $\boldsymbol{x}^*$ vector. As in the proof of Proposition 4, we will consider a reform of size γ in a particular direction (see the proof of Proposition 4 for details of the approach in general). Specifically, it does two things simultaneously. First, it raises the marginal tax rate on x_j exactly at $\boldsymbol{x}^*$. Second, it lowers the marginal tax rate on $x_{j'}$ at $\boldsymbol{x}^*$. In summary, the direction of the form is:

$$\tau\left(\boldsymbol{x}^i\right) = \begin{cases} 0 & \text{if } x_j > x_j^* \text{ and } x_{j'} > x_{j'}^* \\ -1 & \text{if } x_j \leq x_j^* \text{ and } x_{j'} > x_{j'}^* \\ 1 & \text{if } x_j > x_j^* \text{ and } x_{j'} \leq x_{j'}^* \\ 0 & \text{if } x_j \leq x_j^* \text{ and } x_{j'} \leq x_{j'}^* \end{cases} \tag{91}$$

There is no mechanical redistribution here. Using the fact that individuals only respond locally, the impact on welfare is:

$$\int_{\mathcal{I}:y^i=y^*} \left[(1-\psi^*)\mathrm{MTR}^* + \psi^*\widetilde{\mathrm{MTR}}^*\right] \frac{dy^*}{d\gamma} di$$

Since individuals at a y^* are identical, this yields the expression in the proposition. □

Proof of Proposition 6. Since $W_j^* = (1-\psi^*)\,\mathrm{MTR}_j^* + \psi^*\widetilde{\mathrm{MTR}}_j^*$ is assumed to be a constant initially, expression 30 implies that the welfare gain has the same sign as W_j^* if income rises, the opposite sign if income falls, and is unchanged if incomes do not change. Thus, it suffices to focus on the sign of $dy^*/d\gamma$.

First, consider the reform to the marginal tax rate on x_j at y^*. When the policy is adopted, the change to the marginal rate on x_j is:

$$\tau_j\left(\boldsymbol{x}^*\right) d\gamma \equiv \frac{d\tau\left(\boldsymbol{x}^*\right)}{dx_j^*} d\gamma = \frac{d\gamma}{dx_j^*} \tag{92}$$

where dx_j^* is the cross-sectional increase in x_j.

Similarly, the change to the marginal tax rate on $x_{j'}$ is:

$$\tau_{j'}\left(\boldsymbol{x}^*\right) d\gamma \equiv \frac{d\tau\left(\boldsymbol{x}^*\right)}{dx_{j'}^*} d\gamma = \frac{d\gamma}{dx_{j'}^*} \tag{93}$$

We can define the same objects for the perceived tax rates by parameterizing the perceived tax schedule as follows.

$$\widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)=\widetilde{\mathcal{T}}\left(\boldsymbol{x}^i,k\right)+\gamma\tilde{\tau}\left(\boldsymbol{x}^i,k\right) \tag{94}$$

Here, γ is the size of the reform as before, and $\tilde{\tau}\left(\boldsymbol{x}^i,k\right)$ is the direction in which the perceived tax schedule moves as the actual tax schedule is changed.

Using this notation, the change to the perceived marginal tax rate on x_j at y^* is:

$$\widetilde{\tau}_j\left(\boldsymbol{x}^*\right)d\gamma\equiv\frac{d\widetilde{\tau}\left(\boldsymbol{x}^*\right)}{dx_j^*}d\gamma \tag{95}$$

Similarly, the change to rate to the perceived marginal tax rate on $x_{j'}$ is:

$$\widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right)d\gamma\equiv\frac{d\widetilde{\tau}\left(\boldsymbol{x}^*\right)}{dx_{j'}^*}d\gamma \tag{96}$$

Next, let $d\hat{x}_j^*$ be the counterfactual increase in x_j for the type with income y^* for each $j\in\mathcal{J}$, if she were to increase her income slightly. We know that the following condition must hold because the individual is locally indifferent between increasing or decreasing her income.

$$\widetilde{U}_y=\sum_{j\in\mathcal{J}}d\hat{x}_j^*\left[y_j(\boldsymbol{x}^*)-\widetilde{T}_j(\boldsymbol{x}^*\mid k,T)-\chi_j(\boldsymbol{x}^*\mid\theta^*)\right]=0 \tag{97}$$

What is the mechanical impact of the change in relative tax rates on the left hand side of this condition? Since the only change is to marginal tax rates on x_j and $x_{j'}$, it is:

$$\frac{d\widetilde{U}_y}{d\gamma}=d\hat{x}_{j'}^*\times\widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right)-d\hat{x}_j^*\times\widetilde{\tau}_j\left(\boldsymbol{x}^*\right) \tag{98}$$

This now puts us in a position to study the three cases in the proposition.

Case i: Assumptions 2 and 6 both hold

If both marginal tax rate changes are equally salient as defined, then there is no perceived increase in the cross-sectional rate at which the tax burden increases. This means that:

$$dx_{j'}^*\times\widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right)-dx_j^*\times\widetilde{\tau}_j\left(\boldsymbol{x}^*\right)=0 \tag{99}$$

With separability, all types increase make the same decisions given the same level of income. Thus, $d\hat{x}_j^*=dx_j^*$ and $d\hat{x}_{j'}^*=dx_{j'}^*$. As a result, equation 99 implies that equation 98 is zero. Since there is no change to the indifference condition (equation 97), there is no incentive for individuals at y^* to change their income level. Finally—as observed at the start of the proof—with no change in y^*, there is no change in welfare.

Case ii: Assumption 2 holds but not Assumption 6

With separability (Assumption 2), we know that $d\hat{x}_j^* = dx_j^*$ and $d\hat{x}_{j'}^* = dx_{j'}^*$. Thus, the change to the indifference condition is:

$$\frac{d\widetilde{U}_y}{d\gamma} = dx_{j'}^* \times \widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right) - dx_j^* \times \widetilde{\tau}_j\left(\boldsymbol{x}^*\right) \tag{100}$$

However, this is not necessarily zero if the tax changes are not equally salient (a violation of Assumption 6). Specifically, we can write:

$$\frac{d\widetilde{U}_y}{d\gamma} = dx_{j'}^* \times \widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right) - dx_j^* \times \widetilde{\tau}_j\left(\boldsymbol{x}^*\right) \tag{101}$$

$$\propto \left(\frac{\widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right)}{\tau_{j'}\left(\boldsymbol{x}^*\right)} \Big/ \frac{\widetilde{\tau}_j\left(\boldsymbol{x}^*\right)}{\tau_j\left(\boldsymbol{x}^*\right)}\right) dx_{j'}^* \times \tau_{j'}\left(\boldsymbol{x}^*\right) - dx_j^* \times \tau_j\left(\boldsymbol{x}^*\right) \tag{102}$$

Next, recall that there is no overall change in the marginal tax rate at this level of income, since it does not mechanically raise revenue. Thus:

$$dx_{j'}^* \times \tau_{j'}\left(\boldsymbol{x}^*\right) - dx_j^* \times \tau_j\left(\boldsymbol{x}^*\right) = 0 \tag{103}$$

Combining equation 103 with the change to the indifference condition above, $d\widetilde{U}_y/d\gamma$ is positive if:

$$\frac{\widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right)}{\tau_{j'}\left(\boldsymbol{x}^*\right)} > \frac{\widetilde{\tau}_j\left(\boldsymbol{x}^*\right)}{\tau_j\left(\boldsymbol{x}^*\right)} \tag{104}$$

and negative otherwise.

If condition 104 holds, then the individual is no longer indifferent and income increases. If the opposite holds, then income falls. Finally, recall that the welfare gain has the same sign as W_j^* when income rises, and the opposite sign if income falls. Combined, this means that the welfare gain has the same sign as W_j^* if the change to the tax rate on x_j is less salient than the one for $x_{j'}$ (in which case inequality 104 holds) — and the opposite otherwise (in which case the opposite holds).

Case iii: Assumption 6 holds but not Assumption 2

Assumption 6 ensures that the argument in Case (i) holds up until equation 99. However, from there we need to proceed differently. First, we can re-write $d\widetilde{U}_y/d\gamma$ as follows:

$$\frac{d\widetilde{U}_y}{d\gamma} = d\hat{x}^*_{j'} \times \widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right) - d\hat{x}^*_j \times \widetilde{\tau}_j\left(\boldsymbol{x}^*\right) \tag{105}$$

$$\propto \left(\frac{d\hat{x}^*_{j'}}{dx^*_{j'}} \bigg/ \frac{d\hat{x}^*_j}{dx^*_j}\right) dx^*_{j'} \times \widetilde{\tau}_{j'}\left(\boldsymbol{x}^*\right) - dx^*_j \times \widetilde{\tau}_j\left(\boldsymbol{x}^*\right) \tag{106}$$

Combined with equation 99 this is positive if and only if:

$$\frac{d\hat{x}^*_{j'}}{dx^*_{j'}} > \frac{d\hat{x}^*_j}{dx^*_j} \tag{107}$$

and negative otherwise.

If condition 107 holds, then the individual is no longer indifferent and income increases. If the opposite holds, then income falls. Finally, recall that the welfare gain has the same sign as W^*_j when income rises, and the opposite sign if income falls. Combined, this means that the welfare gain has the same sign as W^*_j if higher types earn more of their marginal income via x_j rather than $x_{j'}$, compared to what lower types would choose if they increased their incomes (in which case condition 107 holds)—and the opposite otherwise (in which case the opposite holds). □

Proof of Proposition 7. The proof begins in the same way as the one for Proposition 4. We again start by characterizing the effects of an arbitrary reform of size γ. Such a reform may change each individual's tax burden, their marginal tax rate and government revenue. We first compute these impacts, and then study the implications for welfare.

Tax burden: First, the effect of the reform on individual i's tax burden is:

$$\frac{dT\left(\boldsymbol{x}^i, \gamma\right)}{d\gamma} = \tau\left(\boldsymbol{x}^i\right) + \sum_{j \in J} T_j\left(\boldsymbol{x}^i, \gamma\right) \frac{dx^i_j}{d\gamma} \tag{108}$$

where $T_j\left(\boldsymbol{x}^i, \gamma\right)$ is individual i's initial marginal tax rate with respect to decision x_j.

Marginal tax rates: The effect on this individual's marginal tax rate for decision j is:

$$\frac{dT_j\left(\boldsymbol{x}^i, \gamma\right)}{d\gamma} = \tau_j\left(\boldsymbol{x}^i\right) + \sum_{j \in J} \sum_{j' \in J} T_{jj'}\left(\boldsymbol{x}^i, \gamma\right) \frac{dx^i_j}{d\gamma} \tag{109}$$

where $T_{jj'}\left(\boldsymbol{x}^i, \gamma\right)$ is the effect on this individual's marginal tax rate for decision j when decision j' increases.

As above, these summations take into account the fact that the effect of the reform on tax liability and the marginal tax rate depend on the responses of all decisions, in addition to the direct mechanical effect of the reform itself.

Government revenue: The government budget constraint is:

$$\mathcal{B} = \int_{\mathcal{I}} T\left(\boldsymbol{x}^i, \gamma\right) di \tag{110}$$

so the effect on government revenue is as follows:

$$\frac{d\mathcal{B}}{d\gamma} = \int_{\mathcal{I}} \left[\tau\left(\boldsymbol{x}^i\right) + \sum_{j \in J} T_j\left(\boldsymbol{x}^i, \gamma\right) \frac{dx_j^i}{d\gamma}\right] di \tag{111}$$

Welfare: The direct impact on welfare of the reform is:

$$-\int_{\mathcal{I}} \psi^i \left(\tau\left(\boldsymbol{x}^i\right) + \sum_{j \in J} \left(T_j\left(\boldsymbol{x}^i, \gamma\right) - \widetilde{T}\left(\boldsymbol{x}^i, k, \gamma\right)\right) \frac{dx_j^i}{d\gamma}\right) \tag{112}$$

Just as before, this can be written as:

$$\begin{aligned} \frac{d\mathcal{W}}{d\gamma} = & \int_{\mathcal{I}} \left(1 - \psi^i\right) \tau\left(\boldsymbol{x}^i\right) \\ & + \underbrace{\int_{\mathcal{I}} \left[\sum_j T_j\left(\boldsymbol{x}^i, \gamma\right) \frac{dx_j^i}{d\gamma}\right] di}_{\textbf{Fiscal Externality}} - \underbrace{\int_{\mathcal{I}} \psi^i \left(\sum_{j \in J} \left(T_j\left(\boldsymbol{x}^i, \gamma\right) - \widetilde{T}\left(\boldsymbol{x}^i, k, \gamma\right)\right) \frac{dx_j^i}{d\gamma}\right) di}_{\textbf{Internality}} \end{aligned} \tag{113}$$

Specific perturbation: We next study the optimal amount of redistribution by studying a particular perturbation which raises revenue from the rich and returns that revenue as a lump sum (after netting out any loss due to the fiscal externality). Specifically, let $\tau\left(\boldsymbol{x}\right) = 1$ if $\boldsymbol{x} > \boldsymbol{x}^*$ and zero otherwise. This produces a small change in the marginal tax rate for individuals at y^* without any change to the marginal tax rate elsewhere. Individuals at y^* see their tax rate raised by:

$$\tau_y\left(\boldsymbol{x}^*\right) d\gamma \equiv \frac{\tau\left(\boldsymbol{x}^* + d\boldsymbol{x}\right) - \tau\left(\boldsymbol{x}^*\right)}{dy} d\gamma = \frac{d\gamma}{dy} \tag{114}$$

whereas all other individuals see no change in their (true) marginal tax rates.

In this case, however, we allow individuals to respond to changes in the level of their tax burden as well as their marginal tax rate. This means that individuals with income greater than y^* may re-optimize. Taking into account both types of effect, we can rewrite the change in welfare as:

$$\frac{dW}{d\gamma} = \int_{\mathcal{I}:y^i>y^*} \left(1-\psi^i\right) di + \overbrace{\int_{\mathcal{I}:y^i=y^*} \left[\sum_j T_j\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma}\right] di}^{\textbf{FE (Local)}} \tag{115}$$
$$\underbrace{- \int_{\mathcal{I}:y^i=y^*} \psi^i \left(\sum_{j\in J}\left(T_j\left(\boldsymbol{x}^i,\gamma\right)-\widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)\right)\frac{dx_j^i}{d\gamma}\right) di}_{\textbf{IN (Local)}}$$
$$+ \underbrace{\int_{\mathcal{I}:y^i>y^*} \left[\sum_j T_j\left(\boldsymbol{x}^i,\gamma\right)\frac{dx_j^i}{d\gamma}\right] di}_{\textbf{FE (Non-local)}} - \underbrace{\int_{\mathcal{I}:y^i>y^*} \psi^i \left(\sum_{j\in J}\left(T_j\left(\boldsymbol{x}^i,\gamma\right)-\widetilde{T}\left(\boldsymbol{x}^i,k,\gamma\right)\right)\frac{dx_j^i}{d\gamma}\right) di}_{\textbf{IN (Non-local)}}$$

At a given level of income, we know individuals are identical because $\boldsymbol{x}^i$ increases strictly with θ^i. Combining this fact with the definitions of MTR^i and $\widetilde{\text{MTR}}^i$, we can write the actual and perceived marginal tax rates of individuals at y^* as MTR^* and $\widetilde{\text{MTR}}^*$. Similarly, let MTR^y and $\widetilde{\text{MTR}}^y$ be the average actual and perceived marginal tax rates for those with income y. Finally, let the change in income be $dy^*/d\gamma$ at y^*, and $dy/d\gamma$ at y; and the welfare weight be ψ^* at y^*, and ψ^y at y. Then we can rewrite the welfare change as:

$$\frac{dW}{d\gamma} = \Psi\left(y^*\right) - H\left(y^*\right) - \frac{(1-\psi^*)\text{MTR}^* + \psi^*\widetilde{\text{MTR}}^*}{1-\text{MTR}^*} \times \varepsilon_c^* y^* h\left(y^*\right)$$
$$- \int_{y^*}^{\overline{y}} \frac{(1-\psi^y)\text{MTR}^y + \psi^y\widetilde{\text{MTR}}^y}{1-\text{MTR}^y} \times \eta^y h\left(y\right) dy \tag{116}$$

Setting this welfare change equal to zero yields the equation in the proposition. □

Proof of Proposition 8. We start from social welfare as defined in equation 41 of Section 5, which we restate here for convenience:

$$\boldsymbol{\mathcal{W}}(\beta,T) = \int_i \mathcal{W}\left(y(\boldsymbol{x}^i) - T(\boldsymbol{x}^i) - \chi(\boldsymbol{x}^i \mid \theta^i) - \phi^i(a^i)/\beta\right) di \tag{117}$$

Assumptions 8 and 9 ensure that the individual is initially indifferent between remaining at $a^i = 0$, or slightly increasing it. Assumption 8 then further ensures that a^i increases smoothly as β increases. Assumption 1 means that $\boldsymbol{x}^i$ responds smoothly as a^i changes.

Unlike before, the response is of $\boldsymbol{x}^i$ is uncertain. Nonetheless, we can differentiate $\boldsymbol{\mathcal{W}}(\beta,T)$ with respect to β to obtain the expected first-order impact of a marginal increase in knowledge on social welfare. Note that the direction is important here: β is increased.

While lowering β would reduce the incentive to acquire knowledge, there is no mechanism in the model allowing knowledge to fall.

When we differentiate with respect to β, we obtain:

$$\int_i \mathcal{W}'(U^i) \times \sum_{j\in\mathcal{J}} \left(y^i_j(\boldsymbol{x}^i) - \widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) - \chi_j(\boldsymbol{x}^i \mid \theta^i) \right) E_g\left(\frac{dx^i_j}{d\beta}\right) di \tag{118}$$

where $E_g(dx^i_j/d\beta)$ is the government's expectation of the response of $\boldsymbol{x}$ to a marginal reduction in the cost of acquiring knowledge.

Combining this with the first order condition (equation 3) and using the definition of the welfare weight, we obtain the internality.

$$\lambda \int_i \psi_i \sum_{j\in\mathcal{J}} \left(T_j(\boldsymbol{x}^i) - \widetilde{T}_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) \right) E_g\left(\frac{dx^i_j}{d\beta}\right) di \tag{119}$$

Converting this to government dollars by dividing by λ, and using the definitions of the expected actual and perceived average marginal tax rates, we obtain the internality term (integrated over the population).

$$\int_i \psi_i \left(\mathrm{MTR}^i_e(\boldsymbol{x}^i) - \widetilde{\mathrm{MTR}}^i_e(\boldsymbol{x}^i \mid \boldsymbol{p}_0) \right) E_g\left(\frac{dy^i}{d\beta}\right) di \tag{120}$$

where $\mathrm{MTR}^i_e(\boldsymbol{x}^i)$ and $\widetilde{\mathrm{MTR}}^i_e(\boldsymbol{x}^i \mid \boldsymbol{p}_0)$ are the expected marginal tax rates as defined in the statement of the proposition.

Next, there is the mechanical change in knowledge acquisition costs for each individual (dividing through by λ).

$$\int_i \psi^i \phi^i \left(a^i\right) / \beta^2 di \tag{121}$$

Finally, we have the impact on government revenue via the resource constraint, which is comprised of the direct cost and the fiscal externality.

$$\int_i \left(\sum_{j\in\mathcal{J}} T_j(\boldsymbol{x}^i \mid \boldsymbol{p}_0) E_g\left(\frac{dx^i_j}{d\beta}\right) \right) di - C'(\beta) \tag{122}$$

Using the definition of the average actual marginal tax rate, we obtain:

$$\int_i \mathrm{MTR}^i_e(\boldsymbol{x}^i) E_g\left(\frac{dy^i}{d\beta}\right) di - C'(\beta) \tag{123}$$

Combining this with the internality term and the mechanical costs to individuals, we obtain the change in welfare, $\Delta\boldsymbol{\mathcal{W}}$, stated in Proposition 8 (equation 43). □