

Online Appendix for “Optimal Taxation with Behavioral Agents”

Emmanuel Farhi and Xavier Gabaix

August 2019

Section VIII contains additional results on the paper. Section IX gives much more detail on the Mirrlees model. Section X contains proofs not included in the main paper. The online appendix to the NBER working paper version of this paper (Farhi and Gabaix (2019)) contains further proofs.

VIII Additional Results

VIII.A Complements on Optimal Tax with Heterogeneous Agents

VIII.A.1 Calibration: Optimal Ramsey Tax with Heterogeneous Agents

Here we provide details on the calibration done in Section II.A.

With heterogeneous agents, the misperception is distributed as a 2-point distribution with the following properties:

$$m_i^h = \begin{cases} 1 & \text{with probability } p \\ a & \text{with probability } 1 - p \end{cases}$$

with $a \in [0, 1]$, and

$$\begin{aligned} \mathbb{E}[m_i^h] &= p \times 1 + (1 - p) \times a = 0.25 \\ \mathbb{E}[(m_i^h)^2] &= p \times 1 + (1 - p) \times a^2 = 0.25^2 + 0.13. \end{aligned}$$

These equations are satisfied at $p = .1877$ and $a = .0767$. We then take equation (64), with

$$\begin{aligned} S_i^h &= -\frac{c_i^h \psi_i}{q_i^h} \\ q_i^h &= p_i + m_i^h \tau_i \\ c_i^h &= (q_i^h)^{-\psi_i}. \end{aligned}$$

This yields:

$$\frac{\tau_i^*}{p_i} = \frac{1}{p_i \psi_i \sum_h \pi_h \left[m_i^h \frac{c_i^h}{q_i^h} \left(1 - (1 - m_i^h) \frac{\gamma}{\lambda} \right) \right]}, \quad \left(\sum_h \pi_h c_i^h \right) \left(1 - \frac{\gamma}{\lambda} \right)$$

where $\pi_h \in \{p, (1-p)\}$ is the fraction of agents of each type. Assume values $1 - \frac{\gamma}{\lambda} = \Lambda = 1.25\%$ and $\psi_i = 1$. Then, under the case with heterogeneity ($p = .1877$ and $a = .0767$), we have $\frac{\tau_i^*}{p_i} = 0.0729$, or 7.29%. Under homogeneity with the same average misperception $m_i^h = .25$ for all agents, $\frac{\tau_i^*}{p_i} = 20.3\%$, for a ratio $.203/.0729 = 2.78$. When the taxes are fully salient, so $m_i^h = 1$ for all agents, then the optimal tax is 1.27%, giving a ratio $.0127/.0729 = .174$.

VIII.A.2 Nudges vs. Taxes with Redistributive Concerns

Jointly Optimal Nudges and Taxes We normalize $p = 1$. Agent h has utility $u^h(\mathbf{c}) = c_0^h + \frac{a^h c^h - \frac{1}{2}(c^h)^2}{\Psi}$, so that $\beta^h = \gamma^h$ and $c^h(\tau, \tau^{X,h}) = a^h - \Psi(m^h \tau + \tau^{X,h})$. We investigate the optimal joint policy using both nudges and taxes on goods c . One can show that

$$\frac{\partial^2 L}{\partial \tau \partial \chi} = -\Psi \mathbb{E} \left[(\lambda - \gamma^h (1 - m^h)) \eta^h \right].$$

As a result, if $\gamma^h = \lambda$ so that there are no revenue raising or redistributive motives, then taxes and nudges are substitutes. Taxes and nudges are complements if and only if $\mathbb{E} \left[(\lambda - \gamma^h (1 - m^h)) \eta^h \right] \leq 0$. Nudges and taxes can be complement if social marginal utility of income γ^h and nudgeability η^h are positively correlated. Loosely speaking, if poor agents (with a high γ^h) are highly nudgeable, then taxes and nudges can become complements, because in that case, nudges reduces the consumption of poor nudged agents, thereby improving the redistributive incidence of the tax. We next state the exact values of taxes and nudges, in the case $\gamma^h = \lambda$.

Proposition 17 *Assume $\gamma^h = \lambda$. Then jointly optimal nudges and taxes are given by the following formulas*

$$\tau = \frac{\mathbb{E}[(\eta^h)^2] \mathbb{E}[\tau^{X,h} m^h] - \mathbb{E}[\eta^h m^h] \mathbb{E}[\tau^{X,h} \eta^h]}{\mathbb{E}[(\eta^h)^2] \mathbb{E}[(m^h)^2] - (\mathbb{E}[\eta^h m^h])^2},$$

$$\chi = \frac{\mathbb{E}[\tau^{X,h} \eta^h] \mathbb{E}[(m^h)^2] - \mathbb{E}[\tau^{X,h} m^h] \mathbb{E}[\eta^h m^h]}{\mathbb{E}[(\eta^h)^2] \mathbb{E}[(m^h)^2] - (\mathbb{E}[\eta^h m^h])^2}.$$

The more powerful the nudge is for high-internality agents (the higher is $\mathbb{E}[\tau^{X,h} \eta^h]$, keeping all other moments constant), the more optimal policy relies on the nudge and the less it relies on the tax (the higher is χ , the lower is τ). Symmetrically, if the better perceived is the tax by high-internality people (the higher is $\mathbb{E}[\tau^{X,h} m^h]$), the more optimal policy relies on the tax and the less it relies on the nudge.

The more heterogeneity there is in the perception of taxes (the higher is $\mathbb{E}[(m^h)^2]$, holding all

other moments constant), the less targeted the tax is to the internality/externality, and, as a result, the lower is the optimal tax τ , and under certain conditions, the higher the optimal nudge χ .⁴⁶ Similarly, the more heterogeneity there is in nudgeability (the higher is $\mathbb{E}[(\eta^h)^2]$, holding all other moments constant), then lower is the optimal nudge χ , and, under similar conditions, the higher is the optimal tax τ .

In the general case, we assume heterogenous welfare weights with $\mathbb{E}[\gamma^h] = \lambda$.

Proposition 18 *The optimal tax and nudge satisfy*

$$\tau = \frac{\mathbb{E}[\gamma^h (\eta^h)^2] \mathbb{E}[\lambda \tau^{X,h} m^h - \sigma_{\gamma,c/\Psi}] - \mathbb{E}[\gamma^h \eta^h m^h] \mathbb{E}[\lambda \tau^{X,h} \eta^h]}{\mathbb{E}[\gamma^h (\eta^h)^2] \mathbb{E}[\gamma^h (m^h)^2 - \sigma_{\gamma,m}] - \mathbb{E}[\gamma^h \eta^h m^h] \mathbb{E}[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}]}$$

$$\chi = \frac{\mathbb{E}[\lambda \tau^{X,h} \eta^h] \mathbb{E}[\gamma^h (m^h)^2 - \sigma_{\gamma,m}] - \mathbb{E}[\lambda \tau^{X,h} m^h - \sigma_{\gamma,c/\Psi}] \mathbb{E}[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}]}{\mathbb{E}[\gamma^h (\eta^h)^2] \mathbb{E}[\gamma^h (m^h)^2 - \sigma_{\gamma,m}] - \mathbb{E}[\gamma^h \eta^h m^h] \mathbb{E}[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}]}$$

The proof is in the online appendix to [Farhi and Gabaix \(2019\)](#).

VIII.A.3 Attention as a Good: Proof of Proposition 10

In this subsection we normalize the pre-tax price to 1.

Optimal Taxes with Endogenous Attention: The Case of Small Taxes Given attention $m(\tau)$, the perceived tax is $\tau^s(\tau) = \tau m(\tau)$, and demand is $c(\tau) = y(1 - \psi m(\tau) \tau)$. We assume that attention comes from an optimal cost-benefit analysis:

$$m(\tau) = \arg \max_m -\frac{1}{2} \psi y \tau^2 (1 - m)^2 - g(m).$$

The first term represents the private costs of misunderstanding taxes, $-\frac{1}{2} \psi y (\tau - \tau^s)^2$, while the term $-g(m)$ is the psychic cost of attention, $g(m)$ (see [Gabaix \(2014\)](#)). The planner's problem is $\max_{\tau} L(\tau)$ with

$$L(\tau) = -\frac{1}{2} \psi y m^2(\tau) \tau^2 - A g(m(\tau)) + \Lambda \tau y,$$

where $A = 1$ in the ‘‘optimally allocated attention’’ case and $A = 0$ in the ‘‘no attention cost in welfare’’ case. In the ‘‘fixed attention’’ case, $m(\tau)$ is fixed with $m'(\tau) = 0$, and $g(m) = 0$. The optimal tax satisfies

$$L'(\tau) = -\psi y m(\tau) \tau (m(\tau) + \tau m'(\tau)) - A g'(m(\tau)) m'(\tau) + \Lambda y = 0.$$

⁴⁶The condition is $\mathbb{E}[\tau^{X,h} m^h] \mathbb{E}[(\eta^h)^2] \geq \mathbb{E}[\tau^{X,h} \eta^h] \mathbb{E}[\eta^h m^h]$. It is verified if $\eta^h, m^h, \tau^{X,h}$ are independent.

In the “optimally allocated attention” case, we use the agent’s first order condition $g'(m(\tau)) = \psi y \tau^2 (1 - m(\tau))$ and $A = 1$, and the optimal tax is

$$\tau^{m,*} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m'(\tau)}. \quad (42)$$

In the “no attention cost in welfare case,” $A = 0$, the optimal tax is

$$\tau^{m,0} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m(\tau) m'(\tau)}. \quad (43)$$

When attention is fixed, the optimal tax is

$$\tau^{m,F} = \frac{\Lambda/\psi}{m(\tau)^2}. \quad (44)$$

Proposition 19 *In the interior region where attention has an increasing cost ($\tau m(\tau) m'(\tau) > 0$), the optimal tax is lowest when attention is chosen optimally and its cost is taken into account in welfare; intermediate in the “no attention cost in welfare” case; and largest with fixed attention— $\tau^{m,*} < \tau^{m,0} < \tau^{m,F}$.*

When attention’s cost is taken into account, the planner chooses lower taxes $\tau^{m,*} < \tau^{m,0}$ to minimize both consumption distortions and attention costs.⁴⁷ Plainly, the tax is higher when attention is variable than when attention is fixed—this is basically because demand is more elastic then ($-\frac{p}{c} \frac{\partial c}{\partial \tau} = \psi (m(\tau) + \tau m'(\tau))$).

VIII.B Other Extensions

VIII.B.1 Endogenous Social Cost of Public Funds

Our treatment of the basic Ramsey problem is for a given value of the social cost of public funds λ . Here we briefly show how to account for the potential endogeneity of λ . For simplicity, we confine ourselves to the case of a representative agent.

We assume that the government solves the following planning problem:

$$\max_{\{\tau_i, G\}} \gamma \sum_{i=1}^n \left[\frac{[c_i(\tau_i)]^{1-1/\psi_i} - 1}{1 - 1/\psi_i} - (p_i + \tau_i) c_i(\tau_i) \right] + V(G), \quad (45)$$

subject to the revenue constraint $G = \sum_{i=1}^n \tau_i c_i(\tau_i)$, where $V(G)$ is a concave function of spending on public goods. The social cost of public funds is then given by $\lambda = V'(G)$ where G is computed

⁴⁷The example allows to appreciate the Slutsky matrix with or without constant attention. The Slutsky matrix with constant m has $S_{11|_m}^C = \frac{\partial c(1+\tau, m)}{\partial \tau} = -\psi c m$, while the Slutsky matrix with variable m has $S_{11}^C = \frac{dc(1+\tau, m(\tau))}{d\tau} = -\psi c (m + \tau m'(\tau_1))$.

at the optimum.

Consider the limit of small taxes (i.e., for small $V'(0) - \gamma$). Up to the first order, optimal taxes are still given by equation (17) but now Λ is endogenously given by

$$\Lambda = \frac{V'(0) - \gamma}{\gamma - V''(0) \sum_{i=1}^n \frac{p_i c_i(0)}{\psi_i m_i^2}}. \quad (46)$$

The exogenous Λ case arises when the marginal utility of government spending is constant so that $V''(0) = 0$. The more inattentive agents are to taxes, the less costly it is for the government to raise a given amount of revenues, the more the government spends on public goods, and the higher the taxes that it decides to set.

When there is decreasing marginal utility of government spending ($V''(0) < 0$), the marginal utility of government spending is lower for any given level of government spending, and so the government sets lower taxes, raises less revenues, and spends less on public goods whether or not there is inattention, compared to the case where the marginal utility of government spending is constant. Interestingly, in this case, at the optimum, the social cost of public funds Λ , which is equal to $(V'(G) - \gamma)/V'(G)$, decreases with inattention. This indicates that just like in the exogenous Λ case, the government still raises more revenues and spends more on public goods when agents are more inattentive (but less so than in the exogenous Λ case).

VIII.B.2 Tax Instruments with Differential Salience

We elaborate on a remark we made at the end of Section II.E. As an extreme example, consider again the basic Ramsey example outlined above, and assume that the two tax systems with salience m and m' can be used jointly. Consider the case where there is only one agent and only one (taxed) good. With $m' > m$, we get

$$0 = (\lambda - \gamma) c + [\lambda \tau + \gamma(\bar{\tau}^s - \bar{\tau})] m \mathbf{S}^r, \quad 0 = (\lambda - \gamma) c + [\lambda \tau + \gamma(\bar{\tau}^s - \bar{\tau})] m' \mathbf{S}^r,$$

where $\bar{\tau}^s$ is the total perceived tax arising from the joint perception of the two tax instruments. This requires $\lambda = \gamma$ and with $\bar{\tau}^s = 0$. In other words, the solution is the first best. This is because a planner can replicate a lump sum tax by combining a tax τ with low salience m and a tax $-\tau \frac{m}{m'}$ with high salience $m' > m$, generating tax revenues $\tau \frac{m' - m}{m'}$ per unit of consumption of the taxed good with no associated distortion. This is an extreme result, already derived by Goldin (2015). In general, with more than one agent and heterogeneities in the misperceptions of the two taxes, the first best might not be achievable.

VIII.B.3 A Different Budget Adjustment Rule

The specific formulation of misperception that we have used in this section assumes that the budget adjustments required when agents misperceive taxes are all absorbed by the consumption a good (good 0) with a constant marginal utility. This renders these adjustments relatively painless.

We now explore a variant which increases their costs. We assume that the budget adjustments are concentrated on a “shock absorber” good with a sharply decreasing marginal utility. This increases the distortionary costs of non-salient taxes and reduces optimal taxes in a way that we characterize precisely below. Of course, it is difficult to know a priori which good is the “shock absorber” good (or set of goods) – this is one more place where more empirical evidence is needed to address a behavioral enrichment of the traditional model. It could be some luxury goods (e.g. some restaurant meals), or perhaps more pessimistically investments that can be postponed, e.g. health investments. Our purpose here is only to show how this possibility matters for the results.

When perceived prices q_j^s are different from the true prices q_j , some adjustment is needed for the budget constraint. Let us study a different rule, where a certain good n (“the last good”, imagining a temporal order) bears the brunt of the budget adjustment (it’s a “shock absorber”). This leads to

$$c^{i,s}(\mathbf{q}, \mathbf{q}^s, w) = c^{i,r}(\mathbf{q}^s, w) \text{ for } i \neq n \quad (47)$$

$$c^{n,s}(\mathbf{q}, \mathbf{q}^s, w) = \frac{1}{q_n} \left(w - \sum_{i \neq n} q_i c^{i,s}(\mathbf{q}, \mathbf{q}^s, w) \right). \quad (48)$$

This is: for all goods but the last one, the consumer only pays attention to perceived prices. Only for the last one does she see the budget constraint.⁴⁸ We shall see in the next proposition that we can also write

$$c^{n,s}(\mathbf{q}, \mathbf{q}^s, w) = c^{n,r}(\mathbf{q}^s, w) - \frac{1}{q_n} (\mathbf{q} - \mathbf{q}^s) \cdot \mathbf{c}^r(\mathbf{q}^s, w), \quad (49)$$

i.e. actual consumption of good n is planned consumption $c^{n,r}(\mathbf{q}^s, w)$ minus the adjustment for the surprise $(\mathbf{q} - \mathbf{q}^s) \cdot \mathbf{c}^r(\mathbf{q}^s, w)$ in the actual cost of the goods $i < n$ that have been purchased before good n .

For completeness, we record the Slutsky matrix properties of those rules. (Here we consider the income-compensated matrix \mathbf{S}^C).

Proposition 20 (With the “last good adjusting for the budget” rule) *Consider the model above, with attention m_j to price j . Evaluating at $\mathbf{q}^s = \mathbf{q}$, the marginal propensity to consume out of wealth isn’t changed:*

$$\partial_w c_i^s(\mathbf{q}, \mathbf{q}^s, w) = \partial_w c_i^r(\mathbf{q}, w). \quad (50)$$

⁴⁸Chetty, Looney and Kroft (2009) consider such a rule in a 2-good context. Gabaix (2017) considers such a rule when doing dynamic programming, and the last good is “next period wealth”.

However, the Slutsky matrix S_{ij}^s is changed as follows:

$$S_{ij}^s = S_{ij}^r m_j + \left(\partial_w c_i^r - \frac{1}{q_n} 1_{i=n} \right) (1 - m_j) c^j, \quad (51)$$

where S_{ij}^r is the rational Slutsky matrix.

A Simple Particular Case We next present a simple particular case. Utility is separable, $u(\mathbf{c}) = \sum_{i=0}^n u_i(c_i)$ with $u'_0(c_0) = 1$, $u'_i(c_i) = c_i^{-1/\psi_i}$ for $i = 1, \dots, n-1$ and the “shock absorber” good n has constant marginal utility of $u'_n(c_n) = 1 - \nu < 1$ if $c_n \geq 1$ and $1 + \mu > 1$ if $c_n < 1$.⁴⁹ We call $\mu > 0$ the marginal distortionary cost of budget adjustment. Goods 0 and n cost \$1, and they are untaxed.

The agent chooses his consumption of goods c_0, \dots, c_{n-1} based on the perceived prices $q_i^s = 1 + m_i \tau_i$ and the rest of his money is spent on the last good. Specifically, the demands are as follows. For goods $i = 1, \dots, n-1$, $c_i = (q_i^s)^{-\psi_i}$ (as the consumer solves $u'(c_i) = q_i^s$). The demand for good 0 is $c_0 = w - \sum_{i=1}^{n-1} (q_i^s)^{1-\psi_i} - 1$, as the consumer plans to consume $c_i = (q_i^s)^{-\psi_i}$ for all good $i = 1, \dots, n-1$, and 1 of good n . Once goods 0 through $n-1$ have been purchased, the remaining disposable income for good n is $c_n = w - \sum_{i=0}^{n-1} q_i c_i$.

Then (as derived shortly) the optimal tax on good $i < n$ is as in (16), replacing Λ by

$$\Lambda_i = \frac{\Lambda - (1 - \Lambda)(1 - m_i)\mu}{1 - (1 - \Lambda)(1 - m_i)\mu}. \quad (52)$$

A direct consequence is that the optimal tax τ_i is lower than in the baseline case and is decreasing in μ , particularly for less salient taxes with a small m_i . Indeed, the measure of the social marginal cost of public funds Λ_i is decreasing in the marginal distortionary cost of budget adjustment μ (recall $\Lambda < 1$), coincides with its baseline value of Λ when $\mu = 0$, and is lower than Λ for all $\mu > 0$. Furthermore μ enters the formula through the $\mu(1 - m_i)$ so that these effects are particularly pronounced when attention m_i is low.

We next proceed in greater detail. We take a particular case, which is particularly tractable. There are $n - 2$ goods, and good n is the “shock absorber” good. The price of goods 0 and n is normalized to 1. There’s no tax on goods 0 and n , for simplicity.

Utility is:

$$u(c_0, \dots, c_n) = c_0 + \sum_{i=1}^n u^i(c_i).$$

⁴⁹The level of ν is unimportant provided it is between 0 and 1.

Good 0 has marginal utility of 1, which absorbs income effects, so $v_w = 1$. Hence, $\tau^b = \mathbf{q} - \frac{u_c}{v_w}$ is:

$$\begin{aligned}\tau_0^b &= 0 \\ \tau_i^b &= q_i - q_i^s \text{ for } i = 1, \dots, n-1 \\ \tau_n^b &= 1 - u'_n(c_n)\end{aligned}$$

for $c_i = c_i^r(q_i^s)$ for $1 \leq i < n$ and $c_n = c_n^r + \sum_{i < n} (q_i^s - q_i) c_i$ (from (49)).

Impact on Pigouvian Taxes We revisit our simple model of Section II.B, with an externality on good 1. We have $\lambda = 1$, so that the government's objective function is:

$$L = U(c_1) - (p + \xi) c_1 + u_2(c_2^* - (1 - m) c_1 \tau) + (1 - m) c_1 \tau.$$

i.e. utility from good 1, utility from good 2 (which absorbs the shock $(1 - m) c_1 \tau$), and consumption of good 0 is increased by the lump-sum rebate, which accounts for the last term. The consumer chooses c_1 according to $U'(c_1) = p + m\tau$.

We take utility $U(c_1) = Qc_1 - \frac{c_1^2}{2\Psi}$, so that demand is $c_1 = \Psi(Q - p - m\tau)$. We keep $u'_2(c_2) = 1 + \mu$. We have:

$$\begin{aligned}L'(\tau) &= [p + m\tau - (p + \xi)](-\Psi m) + [-(1 - m)(1 + \mu) + (1 - m)] \frac{d}{d\tau}(c_1 \tau) \\ &= -(m\tau - \xi)\Psi m - (1 - m)\mu(c_1 - \Psi m\tau) \\ &= -(m\tau - \xi)\Psi m - (1 - m)\mu(\Psi(Q - p - 2m\tau)),\end{aligned}$$

which leads to:

$$\begin{aligned}\tau &= \frac{\frac{\xi}{m} - \mu\left(\frac{1-m}{m}\right)(Q-p)}{1 - 2\mu\left(\frac{1-m}{m}\right)} \\ &= \frac{\frac{\xi}{m} - \mu\left(\frac{1-m}{m}\right)\frac{c_1^0}{\Psi}}{1 - 2\mu\left(\frac{1-m}{m}\right)},\end{aligned}$$

where $c_1^0 = \Psi(Q - p)$ is the consumption of good 1 if there is no tax.

Hence, the government doesn't tax the good if: $\xi\Psi < \mu(1 - m)c_1^0$, i.e. if the externality is too small.

IX The Nonlinear Income Tax Problem

Here are the notations we shall use.

$g(z)$: social welfare weight

$h(z)$ (resp. $h^*(z)$): density (resp. virtual density) of earnings z

$H(z)$: cumulative distribution function of earnings
 n : agent's wage, also the index of his type
 $q(z) = R'(z)$: marginal retention rate, locally perceived
 $\mathbf{Q} = (q(z))_{z \geq 0}$: vector of marginal retention rates
 r_0 : tax rebate at 0 income
 $r(z)$: virtual income
 $R(z) = z - T(z)$: retained earnings
 $T(z)$: tax given earnings z
 z : pre-tax earnings
 $\gamma(z)$: marginal social utility of income
 η : income elasticity of earnings
 π : Pareto exponent of the earnings distribution
 ζ^c : compensated elasticity of earnings
 $\zeta_{Q_{z^*}}^c(z)$: compensated elasticity of earnings when the tax rate at z^* changes
 ζ^u : uncompensated elasticity of earnings

IX.A Setup

Agent's Behavior There is a continuum of agents indexed by skill n with density $f(n)$ (we use n rather than h , the conventional index in that literature). Agent n has a utility function $u^n(c, z)$, where c is his one-dimensional consumption, z is his pre-tax income, and $u_z \leq 0$.⁵⁰

The total income tax for income z is $T(z)$, so that disposable income is $R(z) = z - T(z)$. We call $q(z) = R'(z) = 1 - T'(z)$ the local marginal “retention rate”, $\mathbf{Q} = (q(z))_{z \geq 0}$ the ambient vector of all marginal retention rates, and $r_0 = R(0)$ the transfer given by the government to an agent earning zero income. We define the “virtual income” to be $r(z) = R(z) - zq(z)$. Equivalently $R(z) = q(z)z + r(z)$, so that $q(z)$ is the local slope of the budget constraint, and $r(z)$ its intercept.

We use a general behavioral model in a similar spirit to Section I. The primitive is the income function $z^n(q, \mathbf{Q}, r_0, r)$, which depends on the local marginal retention rate q , the ambient vector of all marginal retention rates \mathbf{Q} , $r_0 = R(0)$ the transfer given by the government to an agent earning zero income, and the virtual income r . In the traditional model without behavioral biases we have $z^n(q, \mathbf{Q}, r_0, r) = \arg \max_z u^n(qz + r, z)$, so that z^n does not depend on \mathbf{Q} and r_0 . With behavioral biases, this is no longer true in general. The income function is associated with the indirect utility function $v^n(q, \mathbf{Q}, r_0, r) = u^n(qz + r, z)|_{z=z^n(q, \mathbf{Q}, r_0, r)}$. The earnings $z(n)$ of agent n facing retention schedule $R(z)$ is then the solution of the fixed point problem $z = z^n(q(z), \mathbf{Q}, r_0, r(z))$. His consumption is $c(n) = R(z(n))$ and his utility is $v(n) = u^n(c(n), z(n))$.

⁵⁰If the agent's pre-tax wage is n , L is his labor supply, and utility is $U^n(c, L)$, then $u^n(c, z) = U(c, \frac{z}{n})$. Note that this assumes that the wage is constant (normalized to one). We discuss the impact of relaxing this assumption in Farhi and Gabaix (2019).

Planning Problem The objective of the planner is to design the tax schedule $T(z)$ in order to maximize the following objective function

$$\int_0^\infty W(v(n)) f(n) dn + \lambda \int_0^\infty (z(n) - c(n)) f(n) dn.$$

Like [Saez \(2001\)](#), we normalize $\lambda = 1$. We call $g(n) = W'(v(n)) v_r^n(q(z(n)), \mathbf{Q}, r_0, r(z(n)))$ the marginal utility of income. This is the analogue of β^h in the Ramsey problem of Section I, and we identify agents with their income level $z(n)$ instead of their skill n . Most of the time, we leave implicit the dependence of $n(z)$ on z to avoid cluttering the notations. We now derive a behavioral version of the optimal tax formula in [Saez \(2001\)](#).

IX.B Saez Income Tax Formula with Behavioral Agents

IX.B.1 Elasticity Concepts

Recall that the marginal retention rate is $q(z) = 1 - T'(z)$. Given an income function $z(q, \mathbf{Q}, r_0, r)$, we introduce the following definitions. We define the income elasticity of earnings

$$\eta = qz_r(q, \mathbf{Q}, r_0, r).$$

We also define the uncompensated elasticity of labor (or earnings) supply with respect to the actual marginal retention rate

$$\zeta^u = \frac{q}{z} z_q(q, \mathbf{Q}, r_0, r).$$

Finally, we define the compensated elasticity of labor supply with respect to the actual marginal retention rate

$$\zeta^c = \zeta^u - \eta.$$

We also introduce two other elasticities, which are zero in the traditional model without behavioral biases. We define the compensated elasticity of labor supply at z with respect to the marginal retention rate $q(z^*)$ at a point z^* different from z :

$$\zeta_{Q_{z^*}}^c = \frac{q}{z} z_{Q_{z^*}}(q, \mathbf{Q}, r_0, r).$$

In the main text of the paper, we use the lighter notation $\zeta_{z^*}^c$:

$$\zeta_{z^*}^c \equiv \zeta_{Q_{z^*}}^c. \tag{53}$$

We also define the earnings sensitivity to the lump-sum rebate at zero income⁵¹

$$\zeta_{r_0}^c = \frac{q}{z} z_{r_0} (q, \mathbf{Q}, r_0, r).$$

We shall call $\zeta_{Q_{z^*}}^c$ a “behavioral cross-influence” of the marginal tax rate at z^* on the decision of an agent earning z . In the traditional model with no behavioral biases, $\zeta_{Q_{z^*}}^c = \zeta_{r_0}^c = 0$, not so with behavioral agents.^{52,53}

All these elasticities a priori depend on the agent earnings z . As mentioned above, we leave this dependence implicit most of the time.

Just like in the Ramsey model, we define the “behavioral wedge”

$$\tau^b(q, \mathbf{Q}, r_0, r) = - \frac{qu_c(c, z) + u_z(c, z)}{v_r(q, \mathbf{Q}, r_0, r)} \Big|_{z=z(q, \mathbf{Q}, r_0, r), c=qz+r}.$$

We also define the renormalized behavioral wedge

$$\tilde{\tau}^b(z) = g(z) \tau^b(z).$$

In the traditional model with no behavioral biases, we have $\tau^b(q, \mathbf{Q}, r_0, r) = \tilde{\tau}^b(z) = 0$. But this is no longer true with behavioral agents.

We have the following behavioral version of Roy’s identity:

$$\frac{v_q}{v_w} = z - \frac{\tau^b z}{q} \zeta^c, \quad \frac{v_{Q_{z^*}}}{v_w} = - \frac{\tau^b z}{q} \zeta_{Q_{z^*}}^c. \quad (54)$$

The general model can be particularized to a misperceived utility and a misperceived prices formulation.

Misperceived Prices Model The agent may misperceive the tax schedule, including her marginal tax rate. We call $T^{s,n}(q, \mathbf{Q}, r_0)(z)$ the perceived tax schedule, $R^{s,n}(z) = z - T^{s,n}(q, \mathbf{Q}, r_0)(z)$ the perceived retention schedule, and $q^{s,n}(q, \mathbf{Q}, r_0)(z) = \frac{dR^{s,n}(q, \mathbf{Q}, r_0)(z)}{dz}$ the perceived marginal retention rate. Faced with this tax schedule, the behavior of the agent can be represented by the following problem

$$\operatorname{smax}_{c, z | R^{s,n}(\cdot)} u^n(c, z) \text{ s.t. } c = R(z). \quad (55)$$

⁵¹Formulas would be cleaner without the multiplication by q in those elasticities, but here we follow the public economics tradition.

⁵²For instance, in the misperceived prices model, in general, the marginal tax rate at z^* affects the default tax rate and therefore the perceived tax rate at earnings z .

⁵³In the language of Section I.A, we use income-compensation based notion of elasticity, \mathbf{S}^C , rather than the utility-compensation based notion \mathbf{S}^H .

This formulation implies that the agent's choice (c, z) satisfies $c = R(z)$ and

$$q^{s,n}(z) u_c^n(c, z) + u_z^n(c, z) = 0, \quad (56)$$

instead of the traditional condition $q(z) u_c^n(c, z) + u_z^n(c, z) = 0$. This means that the agent correctly perceives consumption and income (c, z) but misperceives his marginal retention rate $q^{s,n}(z)$. Together with $c = R(z)$, this characterizes the behavior of the agent.⁵⁴

Accordingly, we define $z^n(q, q^s, r)$ to be the solution of $q^{s,n} u_c^n(c, z) + u_z^n(c, z) = 0$ with $c = qz + r$.⁵⁵ The income $z(n)$ of agent n is then the solution of the fixed point equation

$$z = z^n(q(z), q^{n,s}(q, \mathbf{Q}, r_0)(z), r(z)),$$

his consumption is $c(n) = R(z(n))$ and his utility is $v(n) = u^n(c(n), z(n))$.

Summing up, in the misperceived prices model, the primitives are a utility function u and a perception function $q^s(q, \mathbf{Q}, r_0)(z)$. This yields an income function $z(q, q^s, r)$. The general function $z(q, \mathbf{Q}, r_0, r)$ is then $z(q(z'), \mathbf{Q}, r_0, r) = z(q(z'), q^s(q, \mathbf{Q}, r_0)(z'), r)$ for any earnings z' .

One concrete example of misperception is $q^{s,n}(q, \mathbf{Q}, r_0) = q^s(q, \mathbf{Q}, r_0)$ with

$$q^s(q, \mathbf{Q}, r_0)(z) = mq(z) + (1 - m) \left[\alpha q^d(\mathbf{Q}) + (1 - \alpha) \frac{r_0 + \int_0^z q(z') dz'}{z} \right],$$

where $m \in [0, 1]$ is the attention to the true tax (hence retention) rate, $\frac{r_0 + \int_0^z q(z') dz'}{z}$ is the average retention rate (as in [Liebman and Zeckhauser \(2004\)](#)), and $\alpha \in [0, 1]$. The default perceived retention rate might be a weighted average of marginal rates, e.g. $q^d(\mathbf{Q}) = \int q(z) \omega(z) dz$ for some weights $\omega(z)$.

As in the Ramsey case, it is useful to express behavioral elasticities as a function of an agent without behavioral biases. Call $z^r(q^s, r') = \arg \max_z u(q^s z + r', z)$ the earnings of a rational agent facing marginal tax rate q^s and extra non-labor income r' . Then, $z(q, q^s, r) = z^r(q^s, r')$ where r' solves $r' + q^s z^r(q^s, r') = r + q z^r(q^s, r')$. We call $S^r(q^s, r') = \frac{\partial z^r}{\partial q^s}(q^s, r') - \frac{\partial z^r}{\partial r'}(q^s, r') z^r(q^s, r')$ the rational compensated sensitivity of labor supply (it is just a scalar). We also define $\zeta^{cr} = \frac{q S^r}{z}$ as the compensated elasticity of labor supply of the agent if he were rational.

We define $m_{zz} = q_q^s(q, \mathbf{Q}, r_0)(z)$ as the attention to the own marginal retention rate and $m_{zz^*} = q_{Q_{z^*}}^s(q, \mathbf{Q}, r_0)(z)$ as the marginal impact on the perceived marginal retention rate at z of an increase in the marginal retention rate at z^* . Then, we have the following concrete values for the elasticities

⁵⁴This is a sparse max problem with a non-linear budget constraint, which generalizes the sparse max with a linear budget constraint we analyzed in Section II.A. The true constraint is $c = R(z)$, but the perceived constraint is $c = R^{s,n}(q, \mathbf{Q}, r_0)(z)$.

⁵⁵If there are several solutions, we choose the one that yields the greatest utility.

of the general model (the derivation is in [Farhi and Gabaix \(2019\)](#)):

$$\zeta^c = \left(1 - \eta \frac{\tau - \tau^s}{q}\right) \zeta^{cr} m_{zz}, \quad \zeta_{Q_{z^*}}^c = \left(1 - \eta \frac{\tau - \tau^s}{q}\right) \zeta^{cr} m_{zz^*}, \quad (57)$$

$$\tau^b = \frac{\tau - \tau^s}{1 - \eta \frac{\tau - \tau^s}{q}}. \quad (58)$$

If the behavioral agent overestimates the tax rate ($\tau - \tau^s < 0$), the term τ^b is negative. Loosely, we can think of τ^b as indexing an “underperception” of the marginal tax rate. In the traditional model without behavioral biases, $m_{zz^*} = 1_{z=z^*}$, $\tau^s = \tau$ and $\tau^b = 0$.

Misperceived Utility Model In the misperceived utility model, behavior is represented by the maximization of a subjective decision utility $u^s(c, z)$ subject to the budget constraint $c = R(z)$. We then have $\zeta_{Q_{z^*}}^c = 0$, and ζ^c and η are the elasticities associated with decision utility u^s . The behavioral wedge is

$$\tau^b = \frac{\frac{u_c}{u_c^s} u_z^s - u_z}{v_r}. \quad (59)$$

Other Useful Concepts and Notations We next study the impact of the above changes on welfare. Following [Saez \(2001\)](#), we call $h(z)$ the density of agents with earnings z at the optimum and $H(z) = \int_0^z h(z') dz'$. We also introduce the virtual density $h^*(z) = \frac{q(z)}{q(z) - \zeta^c z R''(z)} h(z)$.

We define the social marginal utility of income

$$\gamma(z) = g(z) + \frac{\eta(z)}{1 - T'(z)} \left[\tilde{\tau}^b(z) + (T'(z) - \tilde{\tau}^b(z)) \frac{h^*(z)}{h(z)} \right]. \quad (60)$$

This definition is the analogue of the corresponding definition in the Ramsey model. If the government transfers a lump-sum δK to an agent previously earning z , the objective function of the government increases by $\delta L(z) = (\gamma(z) - 1) \delta K$. The social marginal utility of income $\gamma(z)$ reflects a direct effect $g(z)$ of that transfer to the agent’s welfare, and an indirect effect on labor supply captured—to the leading order as the agent receives δK , his labor supply changes by $\frac{\eta(z)}{1 - T'(z)} \delta K$, which impacts tax revenues by $\frac{\eta(z)}{1 - T'(z)} T'(z) \delta K$ and welfare by $\frac{\eta(z)}{1 - T'(z)} \tilde{\tau}^b(z) \delta K$; the terms featuring $\frac{h^*(z)}{h(z)}$ (in practice often close to 1) capture the fact that the agent’s marginal tax rate changes as the agent adjusts his labor supply, which impacts tax revenues and welfare because of misoptimization.

IX.B.2 Optimal Income Tax Formula

We next present the optimal income tax formula. [Farhi and Gabaix \(2019\)](#) presents the intermediary steps used in the derivation of this formula.

Proposition 21 *Optimal taxes satisfy the following formulas (for all z^*)*

$$\begin{aligned} \frac{T'(z^*) - \tilde{\tau}^b(z^*)}{1 - T'(z^*)} &= \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} (1 - \gamma(z)) \frac{h(z)}{1 - H(z^*)} dz \\ &\quad - \int_0^{\infty} \frac{\zeta_{Q_{z^*}}^c(z)}{\zeta^c(z^*)} \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} \frac{zh^*(z)}{z^* h^*(z^*)} dz. \end{aligned} \quad (61)$$

This formula can also be expressed as a modification of the Saez (2001) formula

$$\begin{aligned} \frac{T'(z^*) - \tilde{\tau}^b(z^*)}{1 - T'(z^*)} &+ \int_0^{\infty} \omega(z^*, z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} dz \\ &= \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} e^{-\int_{z^*}^z \rho(s) ds} \left(1 - g(z) - \eta \frac{\tilde{\tau}^b(z)}{1 - T'(z)} \right) \frac{h(z)}{1 - H(z^*)} dz, \end{aligned} \quad (62)$$

where $\rho(z) = \frac{\eta(z)}{\zeta^c(z)} \frac{1}{z}$ and

$$\omega(z^*, z) = \left(\frac{\zeta_{Q_{z^*}}^c(z)}{\zeta^c(z^*)} - \int_{z'=z^*}^{\infty} e^{-\int_{z^*}^{z'} \rho(s) ds} \rho(z') \frac{\zeta_{Q_{z'}}^c(z)}{\zeta^c(z^*)} dz' \right) \frac{zh^*(z)}{z^* h^*(z^*)}.$$

The first term $\frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} (1 - \gamma(z)) \frac{h(z)}{1 - H(z^*)} dz$ on the right-hand side of the optimal tax formula (61) is a simple reformulation of Saez's formula, using the concept of social marginal utility of income $\gamma(z)$ rather than the marginal social welfare weight $g(z)$ (The link between the two is in equation (60)). The second term $-\frac{1}{z^*} \int_0^{\infty} \frac{\zeta_{Q_{z^*}}^c(z)}{\zeta^c(z^*)} \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} \frac{zh^*(z)}{z^* h^*(z^*)} dz$ on the right-hand side is new and captures a misoptimization effect together with the term $\frac{-\tilde{\tau}^b(z^*)}{1 - T'(z^*)}$ on the left-hand side.

The intuition is as follows. First, suppose for concreteness that $\zeta_{Q_{z^*}}^c(z) > 0$, then increasing the marginal tax rate at z^* leads the agents at another income z to perceive higher taxes on average, which leads them to decrease their labor supply and reduces tax revenues. Ceteris paribus, this consideration pushes towards a lower tax rate, compared to the Saez optimal tax formula. Second, suppose for concreteness that $\tilde{\tau}^b(z) < 0$, then increasing the marginal tax rate at z^* further reduces welfare. This, again, pushes towards a lower tax rate.

The modified Saez formula (62) uses the concept of the social marginal welfare weight $g(z)$ rather than the social marginal utility of income $\gamma(z)$. It is easily obtained from formula (61) using equation (60). When there are no income effects so that $\eta = \rho(z) = 0$, the optimal tax formula (61) and the modified Saez formula (62) are identical. They coincide with the traditional Saez formula when there are no behavioral biases so that $\zeta_{Q_{z^*}}^c(z) = \omega(z^*, z) = \tilde{\tau}^b(z) = 0$. In this case, the left-hand side of (62) is simply $\frac{T'(z^*)}{1 - T'(z^*)}$ so that the formula solves for the optimal marginal tax rate $T'(z^*)$ at z^* .

The formula is expressed in terms of endogenous objects or “sufficient statistics”: social marginal welfare weights $g(z)$, elasticities of substitution $\zeta^c(z)$, income elasticities $\eta(z)$, and income distribution $h(z)$ and $h^*(z)$. With behavioral agents, there are two differences. First, there are two ad-

ditional sufficient statistic, namely the behavioral wedge $\tilde{\tau}^b(z)$ and the behavioral cross-elasticities $\zeta_{Q_{z^*}}^c(z)$. Second, it is not possible to solve out the optimal marginal tax rate in closed form. Instead, the modified Saez formula (62) at different values of z^* form a system of linear equations in the optimal marginal tax rates $T'(z)$ for all z . The formula simplifies greatly in the case where behavioral biases can be represented by a misperceived utility model. Indeed, we then have $\omega(z^*, z) = 0$ and $\tilde{\tau}^b(z) = g(z) \frac{u_c \frac{u_z}{u_c} - u_z}{v_r}$, so that there is no linear system of equations to solve out to recover $T'(z)$.

X Further Proofs and Derivations

X.A General Proofs and Derivations

Proof of Proposition 2 We observe that a tax τ_i modifies the externality as:

$$\frac{d\xi}{d\tau_i} = \sum_h \xi_{c^h} \cdot \left[\mathbf{c}_{q_i}^h(\mathbf{q}, w^h, \xi) + \mathbf{c}_\xi^h \frac{d\xi}{d\tau_i} \right],$$

so $\frac{d\xi}{d\tau_i} = \frac{\sum_h \xi_{c^h} \cdot \mathbf{c}_{q_i}^h}{1 - \sum_h \xi_{c^h} \cdot \mathbf{c}_\xi^h}$. The term $\frac{1}{1 - \sum_h \xi_{c^h} \cdot \mathbf{c}_\xi^h}$ represents the ‘‘multiplier’’ effect of one unit of pollution on consumption, then on more pollution. So, calling $\frac{\partial L}{\partial \tau_i}^{\text{no } \xi}$ the value of $\frac{\partial L}{\partial \tau_i}$ without the externality (that was derived in Proposition 1)

$$\begin{aligned} \frac{\partial L}{\partial \tau_i} - \frac{\partial L}{\partial \tau_i}^{\text{no } \xi} &= \frac{d\xi}{d\tau_i} \left\{ \sum_h W_{v^h} v_w^h \frac{v_\xi^h}{v_w^h} + \lambda \sum_h \boldsymbol{\tau} \cdot \mathbf{c}_\xi^h(\mathbf{q}, w^h, \xi) \right\} = \frac{d\xi}{d\tau_i} \sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda \boldsymbol{\tau} \cdot \mathbf{c}_\xi^h \right] \\ &= \frac{\sum_h \xi_{c^h} \cdot \mathbf{c}_{q_i}^h}{1 - \sum_h \xi_{c^h} \cdot \mathbf{c}_\xi^h} \sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda \boldsymbol{\tau} \cdot \mathbf{c}_\xi^h \right] = \Xi \sum_h \xi_{c^h} \cdot \mathbf{c}_{q_i}^h. \end{aligned}$$

Using Proposition 1,

$$\begin{aligned} \frac{\partial L}{\partial \tau_i} &= \sum_h \left[(\lambda - \gamma^h) c_i^h + \lambda \boldsymbol{\tau} \cdot \mathbf{S}_i^{C,h} - \beta^h \boldsymbol{\tau}^{b,h} \cdot \mathbf{S}_i^{C,h} + \Xi \xi_{c^h} \cdot (-\mathbf{c}_w^h c_i^h + \mathbf{S}_i^{C,h}) \right] \\ &= \sum_h \left[(\lambda - \gamma^h - \Xi \xi_{c^h} \cdot \mathbf{c}_w^h) c_i^h + \lambda \left(\boldsymbol{\tau} + \frac{\Xi}{\lambda} \xi_{c^h} \right) \cdot \mathbf{S}_i^{C,h} - \beta^h \boldsymbol{\tau}^{b,h} \cdot \mathbf{S}_i^{C,h} \right]. \end{aligned}$$

Derivation of (13) We define the internality/externality wedge

$$\boldsymbol{\tau}^{X,h} = \frac{\beta^h}{\lambda} \boldsymbol{\tau}^{I,h} + \boldsymbol{\tau}^{\xi,h}.$$

We have, from (3)

$$\boldsymbol{\tau}^{b,h} = u_{\mathbf{C}}^{s,h}(\mathbf{C}^h) - u_{\mathbf{C}}^h(\mathbf{C}^h) + \mathbf{p} - \mathbf{p} + \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h} = \boldsymbol{\tau}^{I,h} + \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h} = \boldsymbol{\tau}^{I,h} + (\mathbf{I} - \mathbf{M}^h) \boldsymbol{\tau},$$

hence

$$\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \frac{\beta^h}{\lambda} \boldsymbol{\tau}^{b,h} = \boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \frac{\beta^h}{\lambda} (\boldsymbol{\tau}^{I,h} + (\mathbf{I} - \mathbf{M}^h) \boldsymbol{\tau}) = \left[\mathbf{I} - (\mathbf{I} - \mathbf{M}^h) \frac{\beta^h}{\lambda} \right] \boldsymbol{\tau} - \boldsymbol{\tau}^{X,h}.$$

Hence, Proposition 2 implies:

$$\sum_h \left(1 - \frac{\gamma^{\xi,h}}{\lambda} \right) \mathbf{C}^h = - \sum_h (\mathbf{S}^{C,h})' (\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \tilde{\boldsymbol{\tau}}^{b,h}) = - \sum_h \mathbf{M}^{h'} \mathbf{S}^{h,r} \left[\left[\mathbf{I} - (\mathbf{I} - \mathbf{M}^h) \frac{\beta^h}{\lambda} \right] \boldsymbol{\tau} - \boldsymbol{\tau}^{X,h} \right], \quad (63)$$

i.e.

$$\boldsymbol{\tau} = - \left[\sum_h \mathbf{M}^{h'} \mathbf{S}^{r,h} \left(\mathbf{I} - (\mathbf{I} - \mathbf{M}^h) \frac{\gamma^h}{\lambda} \right) \right]^{-1} \sum_h \left[\left(1 - \frac{\gamma^h}{\lambda} \right) \mathbf{C}^h - \mathbf{M}^{h'} \mathbf{S}^{r,h} \boldsymbol{\tau}^{X,h} \right]. \quad (64)$$

Then, in the limit of small taxes (so that $\frac{\gamma^h}{\lambda} \rightarrow 1$), we obtain (13).

Proof of Proposition 4 We start from the Ramsey planning problem in (15). Define

$$L = \gamma \sum_{i=1}^n \left[\frac{(c_i(\tau_i))^{1-1/\psi_i} - 1}{1 - 1/\psi_i} - (p_i + \tau_i) c_i(\tau_i) \right] + \lambda \sum_{i=1}^n \tau_i c_i(\tau_i),$$

where $c_i = (p_i + m_i \tau_i)^{-\psi_i}$. The first-order condition with respect to τ_i is:

$$L_{\tau_i} = \gamma \left[[(c_i(\tau_i))^{-1/\psi_i} - (p_i + \tau_i)] \frac{\partial c_i}{\partial \tau_i} - c_i(\tau_i) \right] + \lambda \left[c_i(\tau_i) + \tau_i \frac{\partial c_i}{\partial \tau_i} \right] = 0.$$

Note that $c_i(\tau_i)^{-1/\psi_i} = p_i + m_i \tau_i$ and $\partial c_i / \partial \tau_i = -\psi_i \frac{c_i}{p_i + m_i \tau_i} m_i$, we can rewrite the FOC as:

$$\begin{aligned} L_{\tau_i} &= \gamma \left[\left(\frac{\lambda}{\gamma} - 1 + m_i \right) \tau_i \frac{-\psi_i c_i(\tau_i) m_i}{p_i + m_i \tau_i} \right] + (\lambda - \gamma) c_i(\tau_i) \\ &= -\lambda \left(\Lambda + \frac{\gamma}{\lambda} m_i \right) \frac{\psi_i \tau_i c_i(\tau_i) m_i}{p_i + m_i \tau_i} + \lambda \Lambda c_i(\tau_i) = 0. \end{aligned}$$

Simplifying gives us:

$$\left(\Lambda + \frac{\gamma}{\lambda} m_i \right) \psi_i \tau_i m_i = \Lambda (p_i + m_i \tau_i)$$

which gives an explicit expression for τ_i :

$$\frac{\tau_i}{p_i} = \frac{\Lambda}{\psi_i m_i} \frac{1}{\Lambda + (1 - \Lambda)m_i - \Lambda/\psi_i} = \frac{\Lambda}{\psi_i m_i^2} \frac{1}{1 + \Lambda \left(\frac{1 - m_i - 1/\psi_i}{m_i} \right)}.$$

Proof of Proposition 6 The government's planning problem is

$$\sum_h U^h(c^h) - (p + \xi^h) c^h. \quad (65)$$

We call $c^{*h} = \arg \max_{c^h} U^h(c^h) - (p + \xi^h) c^h$ the quantity consumed by the agent at the first best.

To make things transparent, we specify

$$U^h(c) = \frac{a^h c - \frac{1}{2} c^2}{\Psi},$$

which using $U_c^h = \frac{a^h - c}{\Psi} = q^s$, implies a demand function $c^h(q^s) = a^h - \Psi q^s$.⁵⁶

After some algebraic manipulations, social welfare compared to the first best can be written as

$$L(\tau) = -\frac{\Psi}{2} \sum_h (m^h \tau - \xi^h)^2. \quad (66)$$

The first best cannot be implemented unless all agents have the same ideal Pigouvian tax, ξ^h/m^h . Heterogeneity in attention creates welfare losses.

Optimal Pigouvian tax. At the optimum, $U_c^h(c^{h*}) = p + \xi^h$. If the agent perceives only $m^h \tau$, his demand is off the ideal c^{h*} (up to second order terms) as:

$$c^h = c^{h*} - \Psi (m^h \tau - \xi^h).$$

This expression is exact in the quadratic functional form above, and otherwise the leading term of a Taylor expansion of a general function, with now the interpretation $\Psi = \frac{1}{U_{cc}^h(c^{h*})}$ then. The social welfare is $L = \sum_h L^h = -\frac{\Psi}{2} \sum_h (m^h \tau - \xi^h)^2$ by (66).

Because $L_\tau = -\Psi \sum_h m^h (m^h \tau - \xi^h)$, the optimal tax is

$$\tau^* = \frac{\sum_h \xi^h m^h}{\sum_h (m^h)^2} = \frac{\mathbb{E}[\xi^h m^h]}{\mathbb{E}[(m^h)^2]}.$$

⁵⁶The expressions in the rest of this section are exact with this quadratic utility specification. For general utility functions, they hold provided that they are understood as the leading order terms in a Taylor expansion around an economy with no heterogeneity.

Let us calculate $V = \mathbb{E} \left[(m^h \tau - \xi^h)^2 \right]$ at this optimum $\tau = \tau^*$,

$$\begin{aligned} V &= \mathbb{E} \left[(m^h)^2 \right] \tau^{*2} - 2\mathbb{E} [m^h \xi^h] \tau^* + \mathbb{E} \left[(\xi^h)^2 \right] \\ &= \mathbb{E} \left[(m^h)^2 \right] \frac{\mathbb{E} [\xi^h m^h]^2}{\mathbb{E} [(m^h)^2]^2} - 2\mathbb{E} [m^h \xi^h] \frac{\mathbb{E} [\xi^h m^h]}{\mathbb{E} [(m^h)^2]} + \mathbb{E} \left[(\xi^h)^2 \right] = -\frac{\mathbb{E} [\xi^h m^h]^2}{\mathbb{E} [(m^h)^2]} + \mathbb{E} \left[(\xi^h)^2 \right] \\ &= \frac{\mathbb{E} \left[(\xi^h)^2 \right] \mathbb{E} \left[(m^h)^2 \right] - \mathbb{E} [\xi^h m^h]^2}{\mathbb{E} [(m^h)^2]}. \end{aligned}$$

hence the welfare loss is: $L = -\frac{1}{2} H \Psi \frac{\mathbb{E} [(\xi^h)^2] \mathbb{E} [(m^h)^2] - (\mathbb{E} [\xi^h m^h])^2}{\mathbb{E} [(m^h)^2]}$.

If there is no tax, the loss is (from equation (66))

$$L^{\text{no tax}} = -\frac{\Psi}{2} \sum_h (m^h \cdot 0 - \xi^h)^2 = -\frac{\Psi}{2} \sum_h (\xi^h)^2 = -\frac{1}{2} H \Psi \mathbb{E} \left[(\xi^h)^2 \right].$$

$$\text{So, } L = L^{\text{no tax}} \frac{\mathbb{E} [(\xi^h)^2] \mathbb{E} [(m^h)^2] - (\mathbb{E} [\xi^h m^h])^2}{\mathbb{E} [(m^h)^2] \mathbb{E} [(\xi^h)^2]}.$$

Optimal quantity mandate. Welfare is $\sum_h [U^h(c^*) - (p + \xi^h) c^*]$. The optimal quantity restriction c^* is characterized by

$$\frac{1}{H} \sum_h U_c^h(c^*) = p + \frac{1}{H} \sum_h \xi^h. \quad (67)$$

The welfare loss compared to the first best, which entails $U_c^h(c^{h*}) = p + \xi^h$ is

$$L^h = \frac{1}{2} U_{cc}^h(c) (c^{h*} - c^*)^2 = -\frac{1}{2} \frac{1}{\Psi} (c^{h*} - c^*)^2.$$

The best consumption satisfies: $L_{c^*}^Q = \sum_h \frac{1}{\Psi} (c^{h*} - c^*) = 0$, i.e. $c^* = \mathbb{E} [c^{h*}]$.

The loss is:

$$L^Q = -\frac{1}{2} \frac{H}{\Psi} \mathbb{E} \left[(c^{h*} - c^*)^2 \right] = -\frac{1}{2} \frac{H}{\Psi} \text{var} (c^{h*}).$$

Proof of Proposition 7 Equation (64) then yields the optimal tax:

$$\boldsymbol{\tau} = (\mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r \mathbf{M}^h])^{-1} \mathbb{E} [\mathbf{M}^{h'}] \mathbf{S}^r \boldsymbol{\tau}^X. \quad (68)$$

with $\boldsymbol{\tau}^X = (\xi_*, 0)'$.

When agents have uniform misperceptions ($\mathbf{M}^h = \mathbf{M}$), the optimal tax is $\boldsymbol{\tau} = \mathbf{M}^{-1} \boldsymbol{\tau}^X$. This implies $\tau_1 = \frac{\xi_*}{m_1} > 0$ and $\tau_2 = 0$. The principle of targeting applies. This is no longer true when misperceptions are not uniform.

We have $(\mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r \mathbf{M}^h])_{ij} = S_{ij}^r \mathbb{E} [m_i^h m_j^h]$ and $(\mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r])_{ij} = \mathbb{E} [m_i^h] S_{ij}^r$. Matrix inversion gives:

$$\tau_2 = \frac{S_{11}^r S_{12}^r (\mathbb{E} [m_1^2] \mathbb{E} [m_2] - \mathbb{E} [m_1 m_2] \mathbb{E} [m_1])}{\det \mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r \mathbf{M}^h]} \xi_*.$$

Because $\mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r \mathbf{M}^h]$ is a dimension 2×2 and has negative roots (there is a good 0, so that \mathbf{S}^r is the block matrix excluding good 0, and has only negative root), $\det \mathbb{E} [\mathbf{M}^{h'} \mathbf{S}^r \mathbf{M}^h] > 0$. The condition in the Proposition is that $\mathbb{E} [m_1^2] \mathbb{E} [m_2] - \mathbb{E} [m_1 m_2] \mathbb{E} [m_1] > 0$. Hence, $\text{sign}(\tau_2) = -\text{sign}(S_{12})$.

The quadratic case simply gives a constant matrix \mathbf{S}^r .

Proof of Proposition 8 We apply (64). Here we have:

$$\begin{aligned} \mathbf{M}^h &= \mathbf{I}, \mathbf{S}^{r,h*} = \begin{pmatrix} -\psi_1 (p_1 + \tau_1)^{-\psi_1 - 1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{C}_2}^{r,h*} \end{pmatrix}, \mathbf{S}^{r,h \neq h*} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{C}_2}^{r,h \neq h*} \end{pmatrix}, \\ c_1^{h*} &= (p_1 + \tau_1)^{-\psi_1}, \boldsymbol{\tau}^{X,h*} = \begin{pmatrix} \frac{\gamma^{h*}}{\lambda} \xi^{h*} \\ \vdots \end{pmatrix}, \boldsymbol{\tau}^{X,h \neq h*} = \begin{pmatrix} 0 \\ \vdots \end{pmatrix}. \end{aligned}$$

We plug the results above into (64). Suppose ρ_p, ρ_r are the portions of agents h^* and agents $h \neq h^*$ in the population with $\rho_p + \rho_r = 1$. Then

$$\begin{aligned} - \left[\sum_h \mathbf{M}^{h'} \mathbf{S}^{r,h} \left(\mathbf{I} - (\mathbf{I} - \mathbf{M}^h) \frac{\gamma^h}{\lambda} \right) \right]^{-1} &= \frac{1}{H} \begin{pmatrix} \rho_p^{-1} \psi_1^{-1} (p_1 + \tau_1)^{\psi_1 + 1} & \mathbf{0} \\ \mathbf{0} & - \left(\rho_p \mathbf{S}_{\mathbf{C}_2}^{r,h*} + \rho_r \mathbf{S}_{\mathbf{C}_2}^{r,h \neq h*} \right)^{-1} \end{pmatrix}, \\ \sum_h \left[\left(1 - \frac{\gamma^h}{\lambda} \right) \mathbf{C}^h - \mathbf{M}^{h'} \mathbf{S}^{r,h} \boldsymbol{\tau}^{X,h} \right] &= H \begin{pmatrix} \rho_p \left[\left(1 - \frac{\gamma^{h*}}{\lambda} \right) (p_1 + \tau_1)^{-\psi_1} + \frac{\gamma^{h*}}{\lambda} \xi^{h*} \psi_1 (p_1 + \tau_1)^{-\psi_1 - 1} \right] \\ \vdots \end{pmatrix} \end{aligned}$$

and get $\tau_1 = \left(1 - \frac{\gamma^{h*}}{\lambda} \right) \frac{p_1 + \tau_1}{\psi_1} + \frac{\gamma^{h*}}{\lambda} \xi^{h*}$, i.e. $\tau_1 = \frac{\frac{\gamma^{h*}}{\lambda} \xi^{h*} + \left(1 - \frac{\gamma^{h*}}{\lambda} \right) \frac{p_1}{\psi_1}}{1 + \left(\frac{\gamma^{h*}}{\lambda} - 1 \right) \frac{1}{\psi_1}}$.

Proof of Proposition 12 For high incomes, $\tilde{\tau}^b$ approaches zero (as high-earnings agents asymptotically accurately perceive their marginal rate, see (23) and (57)), and $\frac{h^*(z)}{h(z)}$ approaches 1.⁵⁷ Hence, from (60), we have

$$\bar{\gamma} = \bar{g} + \bar{\eta}^r \frac{\bar{\tau}}{1 - \bar{\tau}}.$$

We observe that $\frac{1-H(z^*)}{z^* h^*(z^*)} \rightarrow \frac{1}{\pi}$, that $\zeta^c(z^*) = m \zeta^{c,r}(z^*)$, and $\zeta_{z^*}^c(z) = (1-m) \frac{\psi(z^*/z)}{z} \zeta^{c,r}(z)$ (see

⁵⁷Recall that $\frac{h^*(z)}{h(z)} = \frac{q(z)}{q(z) + \zeta^c z T''(z)}$. Calling $a = \lim_{z \rightarrow \infty} z T''(z)$, if we had $a \neq 0$, we'd have $T'(z) \sim a \ln z$, which would diverge for large z . So $a = 0$. Hence, for large z , $\frac{h^*(z)}{h(z)} \rightarrow 1$.

(57)). Taking the large z^* limit in (22) gives:

$$\frac{\bar{\tau}}{1-\bar{\tau}} = \frac{1}{m\bar{\zeta}^{c,r}} \frac{1}{\pi} (1-\bar{\gamma}) - \lim_{z^* \rightarrow \infty} C(z^*),$$

where

$$\begin{aligned} C(z^*) &:= \int_0^\infty \frac{\zeta_{z^*}^c(z)}{\zeta^c(z^*)} \frac{T'(z) - \tilde{\tau}^b(z)}{1-T'(z)} \frac{zh^*(z)}{z^*h^*(z^*)} dz = \frac{1-m}{m} \int_0^\infty \frac{\frac{\psi(z^*/z)}{z} \zeta^{c,r}(z)}{\zeta^{c,r}(z^*)} \frac{T'(z) - \tilde{\tau}^b(z)}{1-T'(z)} \frac{zh^*(z)}{z^*h^*(z^*)} dz \\ &= \frac{1-m}{m} \int_0^\infty \psi(a) \frac{\zeta^{c,r}\left(\frac{z^*}{a}\right) T'\left(\frac{z^*}{a}\right) - \tilde{\tau}^b\left(\frac{z^*}{a}\right) \frac{z^*}{a} h^*\left(\frac{z^*}{a}\right)}{\zeta^{c,r}(z^*) (1-T'\left(\frac{z^*}{a}\right)) z^* h^*(z^*)} \frac{da}{a}, \end{aligned}$$

where we used the change in variables $z = \frac{z^*}{a}$, so $\frac{da}{a} = -\frac{dz}{z}$. Observing that $\frac{T'\left(\frac{z^*}{a}\right) - \tilde{\tau}^b\left(\frac{z^*}{a}\right)}{1-T'\left(\frac{z^*}{a}\right)} \rightarrow \frac{\bar{\tau}}{1-\bar{\tau}}$ and $\frac{\frac{z^*}{a} h^*\left(\frac{z^*}{a}\right)}{z^* h^*(z^*)} \rightarrow a^\pi$, we get

$$\lim_{z^* \rightarrow \infty} C(z^*) = \frac{1-m}{m} \int_0^\infty \psi(a) \frac{\bar{\tau}}{1-\bar{\tau}} a^\pi \frac{da}{a} = \frac{1-m}{m} \frac{\bar{\tau}}{1-\bar{\tau}} A$$

where $A = \int_0^\infty a^{\pi-1} \psi(a) da$. So

$$\frac{\bar{\tau}}{1-\bar{\tau}} = \frac{1}{m\bar{\zeta}^c\pi} \left(1 - \bar{g} - \bar{\eta}^r \frac{\bar{\tau}}{1-\bar{\tau}}\right) - \frac{1-m}{m} \frac{\bar{\tau}}{1-\bar{\tau}} A.$$

Rearranging we get the top marginal rate announced,

$$\bar{\tau} = \frac{1-\bar{g}}{1-\bar{g} + \bar{\eta}^r + \bar{\zeta}^{c,r}\pi(m + (1-m)A)}.$$

- Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. “Salience and Taxation: Theory and Evidence.” *The American Economic Review*, 99(4): 1145–1177.
- Farhi, Emmanuel, and Xavier Gabaix.** 2019. “Optimal Taxation with Behavioral Agents.” *NBER Working Paper No. 21524*.
- Gabaix, Xavier.** 2014. “A Sparsity-Based Model of Bounded Rationality.” *Quarterly Journal of Economics*, 129(4): 1661–1710.
- Gabaix, Xavier.** 2017. “Behavioral macroeconomics via sparse dynamic programming.” *Working Paper*.
- Goldin, Jacob.** 2015. “Optimal Tax Salience.” *Journal of Public Economics*, 131: 115–123.
- Liebman, Jeffrey B., and Richard J. Zeckhauser.** 2004. “Schmeduling.” https://scholar.harvard.edu/files/jeffreyliebman/files/Schmeduling_WorkingPaper.pdf.
- Saez, Emmanuel.** 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies*, 68(1): 205–229.