Optimal Taxation with Behavioral Agents*

Emmanuel Farhi
Harvard, CEPR and NBER

Xavier Gabaix
NYU, CEPR and NBER

August 26, 2015

Abstract

This paper develops a theory of optimal taxation with behavioral agents. We use a general behavioral framework that encompasses a wide range of behavioral biases such as misperceptions, internalities and mental accounting. We revisit the three pillars of optimal taxation: Ramsey (linear commodity taxation to raise revenues and redistribute), Pigou (linear commodity taxation to correct externalities) and Mirrlees (nonlinear income taxation). We show how the canonical optimal tax formulas are modified and lead to a rich set of novel economic insights. We also show how to incorporate nudges in the optimal taxation frameworks, and jointly characterize optimal taxes and nudges. We explore the Diamond-Mirrlees productive efficiency result and the Atkinson-Stiglitz uniform commodity taxation proposition, and find that they are more likely to fail with behavioral agents. (JEL: D03, H21).

*efarhi@fas.harvard.edu, xgabaix@stern.nyu.edu. For very good research assistance we thank Deepal Basak and Chenxi Wang, and for helpful comments we thank seminar participants at Berkeley, BEAM, BRIC, Brown, BU, Chicago, Columbia, NBER, NYU, PSE, Stanford, the UCL conference on behavioral theory, and H. Allcott, R. Chetty, S. Dellavigna, A. Frankel, M. Gentzkow, E. Glaeser, O. Hart, E. Kamenica, L. Kaplow, W. Kopczuk, D. Laibson, U. Malmendier, E. Saez, B. Salanié, J. Schwarzstein, A. Shleifer, and T. Stralezcki. Gabaix thanks INET and the NSF (SES-1325181) for support.

1 Introduction

This paper develops a systematic theory of optimal taxation with behavioral agents. Our framework allows for a wide range of behavioral biases (for example, misperception of taxes, internalities, or mental accounting), structures of demand, externalities, and population heterogeneity, as well as tax instruments. We derive a behavioral version of the three pillars of optimal taxation: Ramsey (1927) (linear commodity taxation to raise revenues and redistribute), Pigou (1920) (linear commodity taxation to correct for externalities), and Mirrlees (1971) (nonlinear income taxation).

Our results take the form of optimal tax formulas that generalize the canonical formulas derived by Diamond (1975), Sandmo (1975), and Saez (2001). Our formulas are expressed in terms of similar sufficient statistics and share a common structure.

The sufficient statistics can be decomposed into two classes: traditional and behavioral. Traditional sufficient statistics, which arise in non-behavioral models, include: social marginal utilities of income and of public funds, compensated demand elasticities, marginal externalities, and equilibrium demands. Behavioral sufficient statistics are misoptimization wedges that arise only in behavioral models. The behavioral tax formulas differ from their traditional counterparts because of the behavioral sufficient statistics and because the presence of behavioral biases shapes the traditional sufficient statistics and, in particular, cross and own price elasticities as well as social marginal utilities of income.

The generality of our framework allows us to unify existing results in behavioral public finance as well as to derive new insights. A non-exhaustive list includes: a modified Ramsey inverse elasticity rule (for a given elasticity, inattention increases the optimal tax, quadratically); a modified optimal Pigouvian tax rule (for a given externality, inattention increases the optimal tax, linearly); a new role for quantity regulation (heterogeneity in attention favors quantity regulation over price regulation); the attractiveness of targeted nudges (which respects freedom of choice for rational agents and limit the tax burden of the poor); a modification of the principle of targeting (in the traditional model, it is optimal to tax the externality-generating good, but not to subsidize substitute goods; in the behavioral model, it is actually optimal to subsidize substitute goods); a uniform commodity taxation result within a "rigid" mental account (i.e. when the total budget on a category of goods is inelastic because of behavioral mental accounting); in the Mirrleesian optimal nonlinear income tax, marginal income tax rates can be negative even with only an intensive labor margin; if the top marginal tax rate is particularly salient and contaminates perceptions of other marginal tax rates, then it should be lower than prescribed in the traditional analysis, and, conversely, if the wealthy overperceive the productivity of effort, top marginal rates are higher than the traditional analysis.

We also revisit two classical results regarding supply elasticities and production efficiency. The first classical result states that optimal tax formulas do not depend directly on supply elasticities if there is a full set of commodity taxes. The second classical result, due to Diamond and Mirrlees (1971), states that under some technical conditions, production efficiency holds at the optimum (so

that, for example, intermediate goods should not be taxed and inputs should be taxed at the same rate in all sectors) if there is a complete set of commodity taxes and if there are constant returns to scale or if profits are fully taxed. We show that both results can fail when agents are behavioral because agents might misperceive taxes. Roughly, a more stringent condition is required, namely, a full set of commodity taxes that agents perceive like prices (in addition perhaps to other commodity taxes which could be perceived differently from prices).

Finally, we show that the celebrated uniform commodity taxation result of Atkinson and Stiglitz (1972) requires more stringent conditions when agents are behavioral.

Relation to the literature We rely on recent progress in behavioral public finance and basic behavioral modelling. We build on earlier behavioral public finance theory.[1] Chetty (2009) and Chetty, Kroft and Looney (CKL, 2009) analyze tax incidence and welfare with misperceiving agents; however, they do not analyze optimal taxation in this context—our paper can thus be viewed as a next logical step after CKL. An emphasis of previous work is on the correction of "internalities," i.e. misoptimization because of self-control or limited foresight, which can lead to optimal "sin taxes" on cigarettes or fats (Gruber and Kőszegi 2001, O'Donoghue and Rabin 2006).

Mullainathan, Schwarzstein and Congdon (2012) offer a rich overview of behavioral public finance. In particular, they derive optimality conditions for linear taxes, in a framework with a binary action and a single good. Baicker, Mullainathan, and Schwartzstein (2015) further develop those ideas in the context of health care. Allcott, Mullainathan and Taubinsky (2014) analyze optimal energy policy when consumers underestimate the cost of gas with two goods (e.g. cars and gas) and two linear tax instruments. The Ramsey and Pigou models in our paper generalize those two analyses by allowing for multiple goods with arbitrary patterns of own and cross elasticities and for multiple tax instruments. We derive a behavioral version of the Ramsey inverse elasticity rule.

Liebman and Zeckhauser (2004) study a Mirrlees framework when agent misperceive the marginal tax rate for the average tax rate. Two recent, independent papers by Gerritsen (2014) and Lockwood (2015) study a Mirrlees problem in a decision vs. experienced model, including a calibration and evidence. Our behavioral Mirrlees framework is general enough to encompass, at a formal level, these models as well as many other relying on alternative behavioral biases.

We also take advantage of recent advances in behavioral modeling of inattention. We use a general framework that reflects previous analyses, including misperceptions and internalities. We rely on the sparse agent of Gabaix (2014) for many illustrations, which builds on the burgeoning literature on inattention (Bordalo, Genaiolli and Shleifer (2013), Caplin and Dean (2015), Chetty, Kroft and Looney (2009), Gabaix and Laibson (2006), Kőszegi and Szeidl (2013), Schwartzstein (2014), Sims (2003), Woodford (2012)). This agent misperceives prices in a way that can be

[1] Numerous studies now document inattention to prices or more broadly consequences of purchases, e.g. Abaluck and Gruber (2011), Allcott and Taubinsky (forthcoming), Allcott and Wozny (2014) (see also Busse, Knittel and Zettelmeyer 2012), Anagol and Kim (2012), Brown, Hossain and Morgan (2010), Chetty (2015), DellaVigna (2009), and Ellison and Ellison (2009).

endogenized to economize on attention (hence the name "sparse"), respects the budget constraint in a way that gives a tractable behavioral version of basic objects of consumer theory, e.g. the Slutsky matrix and Roy's identity. Second, we also use the "decision utility" paradigm, in which the agent maximizes the wrong utility function. We unify those two strands in a general, agnostic framework that can be particularized to various situations.

The rest of the paper is organized as follows. Section 2 gives introduction to the topic, using a minimum of mathematical apparatus to uncover some of the intuitions and economic forces. Section 3 then develops the general theory, with heterogeneous agents, arbitrary utility and decision functions. Section 4 shows a number of examples applying this general theory. Section 5 studies the optimal nonlinear income tax problem. Section 6 studies the impact of endogenous attention and salience as a policy choice. Section 7 revisits Diamond and Mirrlees (1971) and Atkinson and Stiglitz (1972). Section 8 outlines the new empirical measures demanded by the behavioral tax formulas, discusses some of the limitations of our approach, and identifies directions for future work. The online appendix contains more proofs and extensions.

2 The Basic Behavioral Ramsey and Pigou Problems

This section is meant as a simple introduction to the topic of this paper, using a lighter formal apparatus than the rest of the paper. It is in particular designed for the reader who is interested in behavioral insights, but does not wish to pay the fixed cost of acquiring the general notions of traditional canonical taxation theory, as in Section 3. We start with an elementary treatment of the basic Ramsey problem with behavioral agents. We then move to Pigouvian taxation.

The simplicity of the analysis in this section arises from a set of assumptions that we relax later: we mostly abstract from heterogeneity and confine ourselves to the case of a representative agent; we ignore redistributive concerns and focus only on taxes to raise revenues and to correct externalities and internalities; we consider quasi-linear and separable preferences; we study the limit of small taxes; and we restrict ourselves to a simple parametrized set of behavioral biases with misperceptions of taxes (Ramsey) and temptation (Pigou).

Section 3 relaxes all these assumptions and extends the results in this section in the context of our general model, which allows for arbitrary heterogeneity, redistributive motives for taxation, general preferences, large taxes, and general behavioral biases.

2.1 Basic Ramsey Problem: Raising Revenues with Behavioral Agents

2.1.1 The Consumer

We consider a representative agent with utility function $u(\boldsymbol{c})$ and budget constraint $\boldsymbol{q}\cdot\boldsymbol{c}\leq w$. Here $\boldsymbol{c}=(c_0,...,c_n)$ is the consumption vector, $\boldsymbol{p}=(p_0,...,p_n)$ is the before-tax price vector, $\boldsymbol{\tau}=(\tau_0,...,\tau_n)$ is a vector of linear commodity taxes, $\boldsymbol{q}=\boldsymbol{p}+\boldsymbol{\tau}$ is the after-tax price vector, and w is wealth. If

the representative agent were fully rational, he would solve:

$$\max_{\boldsymbol{c}} u(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q} \cdot \boldsymbol{c} \leq w,$$

with resulting demand $\boldsymbol{c}^r(\boldsymbol{q}, w)$ (the superscript r standing for the traditional $r$ational demand).

But the representative agent may not be rational. In the general model in Section 3, we specify a general demand function $\boldsymbol{c}(\boldsymbol{q}, w)$ which reflects arbitrary behavioral biases, but still exhausts the budget $\boldsymbol{q} \cdot \boldsymbol{c}(\boldsymbol{q}, w) = w$. This flexible and general modelling strategy allows us to capture a range of behavioral traits, ranging from temptation to misperception of prices and taxes. In this introductory section, we use a simple specialization of the general model. We assume that behavioral biases take the form of misperception of prices and taxes as captured by the sparse max model developed in Gabaix (2014). Demand depends on both true prices $\boldsymbol{q}$, perceived prices $\boldsymbol{q}^s$ (the superscript s stands for "$s$ubjectively perceived prices" here) and wealth w. It can be represented as

$$\max_{\boldsymbol{c}|\boldsymbol{q}^s} u(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q} \cdot \boldsymbol{c} \leq w.$$

The resulting demand $\boldsymbol{c}^s(\boldsymbol{q}, \boldsymbol{q}^s, w)$ has the following characterization

$$\boldsymbol{c}^s(\boldsymbol{q}, \boldsymbol{q}^s, w) = \boldsymbol{c}^r(\boldsymbol{q}^s, w'),$$

where w' solves $\boldsymbol{q} \cdot \boldsymbol{c}^r(\boldsymbol{q}^s, w') = w$. In other words, the demand of a behavioral agent perceiving prices $\boldsymbol{q}^s$ and with budget w is the demand $\boldsymbol{c}^r(\boldsymbol{q}^s, w')$ of a rational agent facing prices $\boldsymbol{q}^s$ and a different budget w'. The value of w' is chosen to satisfy the true budget constraint, $\boldsymbol{q} \cdot \boldsymbol{c}^r(\boldsymbol{q}^s, w') = w$.[2]

With this formulation, the usual "trade-off" intuition applies in the space of perceived prices: marginal rates of substitution are equal to relative perceived prices $\frac{u'_{c_1}}{u'_{c_2}} = \frac{q_1^s}{q_2^s}$. For example, with quasilinear utility $u(c) = c_0 + U(c_1, ..., c_n)$ and $q_0 = q_0^s = 1$, all wealth effects are absorbed in the demand for good 0, and we can write $c_i^s(\boldsymbol{q}, \boldsymbol{q}^s, w) = c_i^r(\boldsymbol{q}^s)$ for $i \in \{1, ..., n\}$. The demand of the behavioral agent is then simply the demand of a rational agent with perceived prices $\boldsymbol{q}^s$.

Perceived prices $\boldsymbol{q}^s$ are a function of true prices $\boldsymbol{q}$ and wealth w. In the general model in Section 3, we allow for arbitrary perception functions $\boldsymbol{q}^s(\boldsymbol{q}, w)$. The resulting demand function is

$$\boldsymbol{c}(\boldsymbol{q}, w) = \boldsymbol{c}^s(\boldsymbol{q}, \boldsymbol{q}^s(\boldsymbol{q}, w), w),$$

so that demand depends on both perceived prices and true prices, and perceived prices depend on true prices. In this section, we use a simple parametrization and assume that the agent correctly

[2] This adjustment rule is discussed in Chetty, Looney and Kroft (2009), building on Liebman and Zeckhauser (2004) and Gabaix and Laibson (2006); its statement for general constrained maximization problems is systematized in the sparse max.

perceives prices $\boldsymbol{p}$ but perceives taxes to be

$$\tau_i^s(\boldsymbol{\tau}) = m_i \tau_i, \tag{1}$$

where $m_i \in [0,1]$ is an attention parameter. Full attention/rationality corresponds to $m_i = 1$, and full inattention to $m_i = 0$. In the general model we also allow attention m to be endogenous, but we take it to be exogenous in this section. Perceived prices are then given by

$$q_i^s(\boldsymbol{q}, w) = p_i + m_i \tau_i \text{ with } \boldsymbol{q} = \boldsymbol{p} + \boldsymbol{\tau}. \tag{2}$$

2.1.2 Planning Problem

The government must raise revenues using linear commodity taxes $\boldsymbol{\tau}$. We assume that good 0 (e.g. leisure) is untaxed with $\tau_0 = 0$ throughout the paper.[3] In the general model in Section 3, we allow for a general production function. In this section, we assume a simple linear production function. Normalizing $p_0 = 1$, prices are set by the technological resource constraint: $\boldsymbol{p} \cdot \boldsymbol{y} \leq w$, where $w = \boldsymbol{p} \cdot \boldsymbol{e}$ is the value of the initial endowment. The government planning problem is

$$\max_{\boldsymbol{\tau}} \mathcal{L}(\boldsymbol{\tau}),$$

where

$$\mathcal{L}(\boldsymbol{\tau}) = u(\boldsymbol{c}) + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}(\boldsymbol{p} + \boldsymbol{\tau}, w)$$

and $\lambda > 0$ is the marginal utility of public funds.[4]

From now on in this section, we consider a quasi-linear and separable utility

$$u(\boldsymbol{c}) = c_0 + \sum_{i=1}^{n} U^i(c_i).$$

As is common in many introductory expositions of the Ramsey problem, we work in the limit of small taxes (i.e. in the limit of $\Lambda = \lambda - 1$ close to zero). Without loss of generality, we normalize production prices to $p_i = 1$, so that τ_i is both the specific and ad valorem tax on commodity i.

Lemma 2.1 (Approximation of the government objective function). *The government objective function admits the following approximation*

$$\mathcal{L}(\boldsymbol{\tau}) - \mathcal{L}(0) = L(\boldsymbol{\tau}) + o\left(\|\boldsymbol{\tau}\|^2\right) + O\left(\Lambda \|\boldsymbol{\tau}\|^2\right),$$

[3] Otherwise, the taxation problem with a representative agent is trivial. The government can replicate lump-sum taxes with a uniform tax on all commodities.

[4] If the government needs to raise revenues G from taxes, the problem is $\max_{\boldsymbol{\tau}} u(\boldsymbol{c})$ s.t. $\boldsymbol{\tau} \cdot \boldsymbol{c} \geq G$ and $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{p} + \boldsymbol{\tau}, w)$. Then, λ is endogenous and equal to the Lagrange muliplier on the government budget constraint.

where

$$L(\boldsymbol{\tau}) = \frac{-1}{2}\sum_{i=1}^{n}\left(\tau_i^s\right)^2\psi_i y_i + \Lambda\sum_{i=1}^{n}\tau_i y_i. \tag{3}$$

Here τ_i^s is the perceived tax, y_i expenditure on good i at zero taxes, and $\psi_i = -\frac{U^{i\prime}(y_i)}{y_i U^{i\prime\prime}(y_i)}$ is the inverse of the curvature of the utility function U^i for good i at zero taxes or equivalently the demand elasticity for good i of a rational agent.

From now on in this section, we use (3) as the government's objective function in this section. As we shall see, optimal taxes $\boldsymbol{\tau}$ are of order $O(\Lambda)$ so that $L(\boldsymbol{\tau})$ is of order $O(\Lambda^2)$ and captures the leading approximation term in Λ of the government objective function $\mathcal{L}(\boldsymbol{\tau})$. This approximation generalizes the traditional approximation by replacing true taxes τ_i with perceived taxes τ_i^s. The first term $\frac{-1}{2}\sum_{i=1}^{n}\left(\tau_i^s\right)^2\psi_i y_i$ in (3) captures the welfare distortions arising from taxation. Crucially, it is the perceived tax τ_i^s, not the actual tax τ_i, that matters for welfare distortions.[5] The second term, $\Lambda\sum_{i=1}^{n}\tau_i y_i$, captures the net benefit from raising revenues (benefit to the government, minus cost to the agents). There it is the real tax τ_i, not the perceived tax, that matters.

2.1.3 Optimal Taxes

The perceived tax is $\tau_i^s = m_i\tau_i$ with $m_i > 0$. The objective function (3) becomes

$$L(\boldsymbol{\tau}) = \frac{-1}{2}\sum_{i} m_i^2\tau_i^2\psi_i y_i + \Lambda\sum_{i}\tau_i y_i.$$

Maximization over τ_i yields the following result.

Proposition 2.1 (Modified Ramsey inverse elasticity rule) *In the basic Ramsey problem with misperceptions, optimal taxes follow an inverse-elasticity rule modified by inattention*

$$\tau_i = \frac{\Lambda}{m_i^2\psi_i}, \qquad \tau_i^s = \frac{\Lambda}{m_i\psi_i}. \tag{4}$$

This Proposition generalizes the traditional Ramsey inverse elasticity rule, which prescribes that the optimal tax should be $\tau_i^R = \frac{\Lambda}{\psi_i}$ (this traditional rule is recovered when consumers are fully attentive so that $m_i = 1$ for all i). Optimal taxes are higher at $\tau_i = \frac{\Lambda}{m_i^2\psi_i}$ than they would be in the traditional Ramsey solution. Loosely speaking, this is because inattention makes agents less elastic. Given partial attention $m_i \leq 1$, the effective elasticity of the demand for good i is $m_i\psi_i$, rather than the parametric elasticity ψ_i. In the spirit of the traditional Ramsey formula, lower elasticity leads to higher optimal taxes.[6]

[5] This effect is anticipated in Chetty, Looney and Kroft (2009).

[6] Finkelstein (2009) finds evidence for this effect. When highway tolls are paid automatically thus are less salient, people are less elastic to them, and government react by increase the toll (i.e., the tax rate).

However, a naive application of the Ramsey rule would lead to the erroneous conclusion that $\tau_i = \frac{\Lambda}{m_i \psi_i}$ rather than $\tau_i = \frac{\Lambda}{m_i^2 \psi_i}$. To gain intuition for this discrepancy, consider the effect of a marginal increase in τ_i. The marginal benefit in terms of increased tax revenues is Λy_i. The marginal cost in terms of increased distortions is $\tau_i^s m_i \psi_i y_i$. At the optimum, the marginal cost and the marginal benefit are equalized. The result is that $\tau_i^s = \frac{\Lambda}{m_i \psi_i}$, i.e. it is the perceived tax τ_i^s that is inversely related to the effective elasticity $m_i \psi_i$. This in turns implies $\tau_i = \frac{\tau_i^s}{m_i} = \frac{\Lambda}{m_i^2 \psi_i}$. [7]

The upshot of this analysis is that optimal taxes τ_i increase relatively fast with inattention m_i.[8] Formally, taxes increase quadratically with inattention, so that partial attention m_i leads to a multiplication of the traditional tax by $\frac{1}{m_i^2}$.

2.2 Basic Pigou Problem: Externalities, Internalities and Inattention

2.2.1 Optimal Corrective Taxes

We continue to assume a quasilinear utility function. We assume that there is only one taxed good $n = 1$. The representative agent maximizes $u(c_0, c) = c_0 + U(c)$ subject to $c_0 + pc \leq w$. Here c stands for the consumption the good 1 (we could call it c_1, but expressions are cleaner by calling it c). If the representative agent were rational, he would solve

$$\max_c U(c) - pc.$$

However, there is a negative externality that depends on the aggregate consumption of good 1 (think for example of second-hand smoke), so that total utility is $c_0 + U(c) - \xi_* c$. Alternatively, ξ_* could be an internality: a divergence between decision utility $c_0 + U(c)$ and experienced utility $c_0 + U(c) - \xi_* c$. This would capture the idea that good 1 is tempting and has extra unperceived negative effects $\xi_* c$ (e.g. a heart attack). The analysis is identical in both cases.[9]

To focus on the corrective role of taxes, we assume that $\Lambda = 0$ and that the government can rebate tax revenues lump-sum to consumers. The objective function of the government is therefore

$$U(c) - (p + \xi_*) c \ . \tag{5}$$

[7] In the more general case with perceptions $\tau_i^s = m_i \tau_i + (1 - m_i) \tau^d$, for some "default tax" function $\tau^d(\boldsymbol{\tau})$ (which could be the average tax, for instance), we can write $L(\boldsymbol{\tau}, \tau^d) = \frac{-1}{2} \sum_i \left(m_i \tau_i + (1 - m_i) \tau^d \right)^2 \psi_i y_i + \Lambda \sum_i \tau_i y_i$. We have $\frac{dL(\boldsymbol{\tau})}{d\tau_i} = -\tau_i^s m_i \psi_i y_i + \Lambda y_i + \frac{\partial L(\boldsymbol{\tau}, \tau^d)}{\partial \tau^d} \frac{\partial \tau^d(\boldsymbol{\tau})}{\partial \tau_i}$ with $\frac{\partial L(\boldsymbol{\tau}, \tau^d)}{\partial \tau^d} = -\sum_j \tau_j^s (1 - m_j) \psi_j y_j$. The optimal tax is then given by

$$\tau_i = \frac{\Lambda}{m_i^2 \psi_i} + \frac{\frac{\partial L}{\partial \tau^d} \frac{\partial \tau^d}{\partial \tau_i}}{m_i^2 \psi_i y_i} - \frac{1 - m_i}{m_i} \tau^d.$$

Since $\frac{\partial L}{\partial \tau^d} \leq 0$, the optimal tax on a good i is lower when the tax on this good raises the default tax (the term $\frac{\partial \tau^d}{\partial \tau_i}$).

[8] This result relies on the adjustment for error being absorbed by the linear good, good 0. See Proposition 4.8 for an example where the error is absorbed by the good itself.

[9] In this simple example, the internality is analyzed in a similar way to an externality. As we shall see, it is not true in general that the behavioral biases that concern us can be modelled as externalities.

To attempt to correct the externality/internality, the government can set a tax τ. If the agent correctly perceived taxes, he would solve

$$\max_c U(c) - (p+\tau)c.$$

The optimal tax is $\tau = \xi_*$. It ensures that the agent maximizes the same objective function as that of the government. This is the classic Pigouvian prescription: the optimal tax makes the agent exactly internalize the externality/internality. Now consider an agent who only perceives a fraction m of the tax. Then he solves

$$\max_c U(c) - (p+m\tau)c. \tag{6}$$

The optimal Pigouvian corrective tax required to ensure that agents correctly internalize the externality/internality is now $\tau = \frac{\xi_*}{m}$. We record this simple result.

Proposition 2.2 (Modified Pigou formula) *In the basic Pigou problem with misperceptions, the optimal Pigouvian corrective tax is modified by inattention according to* $\tau = \frac{\xi_*}{m}$.

Suppose for concreteness that a good has a negative externality of \$1. With rational agents, it should be taxed by exactly \$1. This is the "dollar-for-dollar" principle of traditional Pigouvian taxation. Accounting for misperception leads to a relaxation of this principle. Indeed, suppose that agents perceive only half of the tax. Then, the good should be taxed by \$2, so that agents perceive a tax of \$1.

It may be contrasted to the modified optimal Ramsey tax (Proposition 2.1), which had $\tau_i = \frac{\Lambda}{\psi_i}\frac{1}{m_i^2}$. Partial attention m_i leads to a multiplication of the traditional tax by $\frac{1}{m_i}$ in the Pigou case and by $\frac{1}{m_i^2}$ in the Ramsey case. [10]

If different consumers have heterogenous perceptions, then Proposition 2.2 suggests that no uniform tax can perfectly correct all of them. Hence, heterogeneity in attention prevents the implementation of the first best. We now explore this issue more thoroughly.

2.2.2 Inattention and a Rationale for Quantity Regulation

We now assume that there are several consumers, indexed by $h = 1...H$. Agent h maximizes $u^h\left(c_0^h, c^h\right) = c_0^h + U^h\left(c^h\right)$. The associated externality/internality is $\xi^h c^h$. He pays an attention m^h to the tax so that perceived taxes are $\tau_h^s = m^h\tau$. The government is utilitarian, so that the government planning problem is

$$\sum_h U^h\left(c^h\right) - \left(p+\xi^h\right)c^h. \tag{7}$$

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

[10] The fact that the tax should be higher when it is underperceived is qualitatively anticipated in Mullainathan, Schwartzstein and Cogdon (2012), but without the specific $\frac{1}{m}$ factor.

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 $c^h(q^s) = a^h - \Psi q^s$.[11]

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

$$L(\tau) = -\frac{\Psi}{2}\sum_h \left(m^h\tau - \xi^h\right)^2. \tag{8}$$

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.

This opens up a potential role for quantity regulations. Suppose the government imposes a uniform quantity restriction, mandating $c^h = c^*$. For instance, banning the good corresponds to setting $c^* = 0$. The following proposition compares optimal Pigouvian regulation and optimal quantity regulation. We consider a situation where the planner implements either an optimal Pigouvian tax, or an optimal quantity regulation, but not both policies.

Proposition 2.3 (Pigouvian tax vs Quantity regulation) *Consider a Pigouvian tax or a quantity restriction in the basic Pigou problem with misperceptions and heterogeneity. The optimal Pigouvian tax is* $\tau^* = \frac{\mathbb{E}\left[\xi^h m^h\right]}{\mathbb{E}\left[m^{h\,2}\right]}$. *The optimal quantity restriction* $c^* = \mathbb{E}\left[c^{h*}\right]$. *Quantity restrictions are superior to corrective taxes if and only if*

$$\frac{1}{2\Psi}var\left(c^{h*}\right) < \Psi\frac{\mathbb{E}\left[\xi^{h\,2}\right]\mathbb{E}\left[m^{h\,2}\right] - \left(\mathbb{E}\left[\xi^h m^h\right]\right)^2}{2E\left[m^{h\,2}\right]}. \tag{9}$$

where the left-hand side is the welfare loss under optimal quantity regulation, and the right-hand side the welfare loss under optimal Pigouvian taxation.

Several insights can be gleaned from (9). First, quantity restrictions tend to dominate if attention m^h is very heterogeneous. Taxes are better if preferences are very heterogeneous (the optimal quantities c^{h*} are very heterogenous). Second, when the demand elasticity is low (low Ψ), quantity restrictions are worse. This is because agents suffer more from a given deviation from their optimal quantity, an effect reminiscent of Weitzman (1974). Third, for small enough externalities/internalities, taxes are better than quantity restrictions. Indeed, there are only second order losses from the externalities/internalities ($\mathbb{E}\left[\xi^{h\,2}\right]$), while the quantity restriction discretely lowers

[11] The 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.

welfare (to the zeroth order).[12] A natural instrument in this context is a nudge, which we develop in detail below in the context of the general model in Section 3.

We have concluded our introductory tour of the most elementary impacts of bounded rationality on taxation. We now move on to a more general analysis.

3 Optimal Linear Commodity Taxation

In this section, we introduce our general model of behavioral biases. We then describe how the basic results of price theory are modified in the presence of behavioral biases. Armed with these results, we then analyze the problem of optimal linear commodity taxation without externalities (Ramsey) where the objective of the government is to raise revenues and redistribute, and with externalities (Pigou) where an additional objective is to correct externalities. We also propose a model of nudges. We show how to incorporate nudges in the optimal taxation framework and characterize the joint optimal use of taxes and nudges. This analysis is performed at a general and rather abstract level. In the next section, we will derive a number of concrete results in a series of simple examples. These examples are simple particularizations of the general model and results.

3.1 Price Theory with Behavioral Agents

Consider a consumer. Our primitive is a demand function $\boldsymbol{c}(\boldsymbol{q}, w)$ where $\boldsymbol{q}$ is the price vector and w is the budget. The demand function incorporates all the behavioral biases that the agent might be subject to (internalities, misperceptions, etc.). The only restriction that we impose on this demand function is that it exhausts the agent's budget so that $\boldsymbol{q} \cdot \boldsymbol{c}(\boldsymbol{q}, w) = w$. We evaluate the welfare of this agent according to a utility function $u(\boldsymbol{c})$. This utility function represents the agent's true or "experienced" utility, with a resulting indirect utility function given by $v(\boldsymbol{q}, w) = u(\boldsymbol{c}(\boldsymbol{q}, w))$. Crucially, the demand function $\boldsymbol{c}(\boldsymbol{q}, w)$ is not assumed to result from the maximization of the utility function $u(\boldsymbol{c})$ subject to the budget constraint $\boldsymbol{q} \cdot \boldsymbol{c} = w$.

This formulation imposes another implicit restriction, namely that producer prices $\boldsymbol{p}$ and taxes $\boldsymbol{\tau}$ matter to the consumer only through $\boldsymbol{q} = \boldsymbol{p} + \boldsymbol{\tau}$. We relax this assumption in Sections 6.2 and 7.1 where we consider a demand function of the form $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{\tau}, w)$. Note however that this distinction is immaterial if producer prices $\boldsymbol{p}$ are exogenously fixed, an assumption which we maintain throughout this section, but which we relax in Section 7.1.

We now introduce the basic objects of price theory and explain how their relations are modified in the presence of behavioral biases. We only highlight the main results. We refer the reader to Appendix 11 for the detailed derivations.

[12] This third point is specific to a quantity mandate (which has no free disposal). It would not hold with a quantity ceiling (which has free disposal).

We define two different Slustky matrices. First, there is the "income-compensated" Slutsky matrix

$$\boldsymbol{S}_j^C(\boldsymbol{q},w) = \boldsymbol{c}_{q_j}(\boldsymbol{q},w) + \boldsymbol{c}_w(\boldsymbol{q},w)c_j(\boldsymbol{q},w).$$

Second, there is the "utility-compensated" Slutsky matrix

$$\boldsymbol{S}_j^H(\boldsymbol{q},w) = \boldsymbol{c}_{q_j}(\boldsymbol{q},w) - \boldsymbol{c}_w(\boldsymbol{q},w)\frac{v_{q_j}(\boldsymbol{q},w)}{v_w(\boldsymbol{q},w)}.$$

The utility-compensated Slutsky matrix $\boldsymbol{S}_j^H(\boldsymbol{q},w)$ can also be computed using the expenditure function $e(\boldsymbol{q},\widehat{u}) = \min_w w$ s.t. $v(\boldsymbol{q},w) \geq \widehat{u}$ and the corresponding Hicksian demand function $\boldsymbol{h}(\boldsymbol{q},\widehat{u}) = \boldsymbol{c}(\boldsymbol{q},e(\boldsymbol{q},\widehat{u}))$. Indeed, we have

$$\boldsymbol{S}_j^H(\boldsymbol{q},v(\boldsymbol{q},w)) = \boldsymbol{h}_{q_j}(\boldsymbol{q},v(\boldsymbol{q},w)).$$

The two Slustky matrices $\boldsymbol{S}_j^C(\boldsymbol{q},w)$ and $\boldsymbol{S}_j^H(\boldsymbol{q},w)$ correspond to two different ways of decomposing the marginal consumption change $c_{q_j}(\boldsymbol{q},w)dq_j$ resulting from a marginal price change dq_j into an income effect and a substitution effect depending on whether the substitution effect is assumed to be computed at constant income or at constant utility. In the first case, the income effect is $-\boldsymbol{c}_w(\boldsymbol{q},w)c_j(\boldsymbol{q},w)dq_j$ and the substitution effect is $\boldsymbol{S}_j^C(\boldsymbol{q},w)dq_j$. In the second case, the income effect is $\boldsymbol{c}_w(\boldsymbol{q},w)\frac{v_{q_j}(\boldsymbol{q},w)}{v_w(\boldsymbol{q},w)}dq_j$ and the substitution effect is $\boldsymbol{S}_j^H(\boldsymbol{q},w)\,dq_j$.

In the traditional model with no behavioral biases, income-compensated marginal prices changes preserve utility since by Roy's identity $v_{q_j}(\boldsymbol{q},w)\,dq_j + v_w(\boldsymbol{q},w)\,c_j(\boldsymbol{q},w)dq_j = 0$. As a result, $\boldsymbol{S}_j^C(\boldsymbol{q},w) = \boldsymbol{S}_j^H(\boldsymbol{q},w)$ and the two decompositions coincide. But with behavioral biases, income-compensated marginal prices changes do not preserve utility since, as we shall see below, Roy's identity fails. As a result, $\boldsymbol{S}_j^C(\boldsymbol{q},w) \neq \boldsymbol{S}_j^H(\boldsymbol{q},w)$ and the two decompositions differ.

With behavioral biases, two other properties of Slutsky matrices are modified. First, in general, the Slutsky matrices $\boldsymbol{S}^C(\boldsymbol{q},w)$ and $\boldsymbol{S}^H(\boldsymbol{q},w)$ are not symmetric. Second, in general, the vectors $\boldsymbol{S}^C(\boldsymbol{q},w)\cdot\boldsymbol{q}$, $\boldsymbol{S}^H(\boldsymbol{q},w)\cdot\boldsymbol{q}$ and $\boldsymbol{q}\cdot\boldsymbol{S}^H(\boldsymbol{q},w)$ are not necessarily equal to 0.[13]

To see how Roy's identity is modified in the presence of behavioral biases, it is useful to first define the misoptimization wedge

$$\boldsymbol{\tau}^b(\boldsymbol{q},w) = \boldsymbol{q} - \frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{q},w))}{v_w(\boldsymbol{q},w)}. \tag{10}$$

[13] We treat the case of $\boldsymbol{S}_j^H(\boldsymbol{q},w)$. In the traditional model with no behavioral biases, we can rewrite the planning problem defining the expenditure function as $e(\boldsymbol{q},\widehat{u}) = \min_{\boldsymbol{c}}\boldsymbol{q}\cdot\boldsymbol{c}$ s.t. $u(\boldsymbol{c}) \geq \widehat{u}$ and apply the envelope theorem to get $\boldsymbol{h}(\boldsymbol{q},\widehat{u}) = e_{\boldsymbol{q}}(\boldsymbol{q},\widehat{u})$. This in turn implies that $\boldsymbol{S}_{ij}^H(\boldsymbol{q},v(\boldsymbol{q},w)) = e_{q_iq_j}(\boldsymbol{q},\widehat{u})$ is necessarily symmetric. And together with the homogeneity of degree 0 of the Hicksian demand function $\boldsymbol{h}(\boldsymbol{q},\widehat{u})$, this implies that $\boldsymbol{S}^H(\boldsymbol{q},w)\cdot\boldsymbol{q} = \boldsymbol{q}\cdot\boldsymbol{S}^H(\boldsymbol{q},w) = 0$. With behavioral biases, this rewriting of the planning problem defining the expenditure function is invalid. As a result, in general $\boldsymbol{S}_{ij}^H(\boldsymbol{q},v(\boldsymbol{q},w)) \neq e_{q_iq_j}(\boldsymbol{q},\widehat{u})$ so that we cannot conclude that $\boldsymbol{S}^H(\boldsymbol{q},w)$ is symmetric. See below for an explicit example using the misperception model.

where b refers to a wedge to due $b$ehavioral biases. In the traditional model without behavioral biases, $\boldsymbol{\tau}^b(\boldsymbol{q}, w) = 0$. The wedge $\boldsymbol{\tau}^b(\boldsymbol{q}, w)$ is an important sufficient statistic for behavioral biases.

Armed with the definition of $\boldsymbol{\tau}^b(\boldsymbol{q}, w)$, we can now state the behavioral version of Roy's identity

$$\frac{v_{q_j}(\boldsymbol{q}, w)}{v_w(\boldsymbol{q}, w)} = -\boldsymbol{c}_j(\boldsymbol{q}, w) - \boldsymbol{\tau}^b(\boldsymbol{q}, w) \cdot \boldsymbol{S}_j^C(\boldsymbol{q}, w), \tag{11}$$

where the term $\boldsymbol{\tau}^b(\boldsymbol{q}, w) \cdot \boldsymbol{S}_j^C(\boldsymbol{q}, w)$ is a discrepancy term that arises from a failure of the envelope theorem because agents do not fully optimize. The intuition for equation (11) is the following: the impact on welfare $v_{q_j}(\boldsymbol{q}, w)\, dq_j = u'(\boldsymbol{c})\boldsymbol{c}_{q_j}(\boldsymbol{q}, w)dq_j$ of a change dq_j in the price of commodity j can be decomposed into an income effect $-u'(\boldsymbol{c})\boldsymbol{c}_w(\boldsymbol{q}, w)c_j(\boldsymbol{q}, w)dq_j = -v_w(\boldsymbol{q}, w)\, c_j(\boldsymbol{q}, w)dq_j$ and a substitution effect $u'(\boldsymbol{c}) \cdot \boldsymbol{S}_j^C(\boldsymbol{q}, w)dq_j$. In the traditional model with no behavioral biases, the income-compensated price change that underlies the substitution effect does not lead to any change in welfare—an application of the envelope theorem. The traditional version of Roy's identity follows. As we have already argued above, with behavioral biases, income-compensated price changes lead to changes in welfare—a failure of the envelope theorem. The behavioral version of Roy's identity accounts for the associated welfare effects.

We now present two useful concrete particularizations of the general model: the decision vs. experienced utility model and the misperception model.

Decision vs. experienced utility model We start with the decision vs. experienced utility model, in which the demand function could arise from a "decision utility" $u^s(\boldsymbol{c})$ (the $s$ubjectively perceived utility), so that

$$\boldsymbol{c}(\boldsymbol{q}, w) = \arg\max_{\boldsymbol{c}} u^s(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q} \cdot \boldsymbol{c} \leq w.$$

However, the true "experienced" utility remains $u(\boldsymbol{c})$ which can be different from $u^s(\boldsymbol{c})$. In this case, the misoptimization wedge is simply given by the wedge between the decision and experienced marginal utilities

$$\boldsymbol{\tau}^b(\boldsymbol{q}, w) = \frac{u_c^s(\boldsymbol{q}, w)}{v_w^s(\boldsymbol{q}, w)} - \frac{u_c(\boldsymbol{q}, w)}{v_w(\boldsymbol{q}, w)}.$$

The income-compensated Slutsky matrix $\boldsymbol{S}_j^C(\boldsymbol{q}, w)$ is the Slutsky matrix of an agent with utility $u^s(\boldsymbol{c})$. It is different from the utility-compensated Slutsky matrix $\boldsymbol{S}_j^H(\boldsymbol{q}, w)$ since the underlying compensation is for experienced utility $u(\boldsymbol{c})$ rather than decision utility $u^s(\boldsymbol{c})$.

As an example, consider the case of two goods $\boldsymbol{c} = (c_0, c_1)$, quasilinear utilities $u^s(\boldsymbol{c}) = c_0 + U^s(c_1)$ and $u(\boldsymbol{c}) = c_0 + U(c_1)$ with $U(c_1) = U^s(c_1) - \xi c_1$ and $\xi > 0$. The consumption c_1 of good 1 entails a negative internality $-\xi c_1$. For instance, good 1 could be ice creams, $U^s(c_1)$ would then represent the immediate pleasure from eating the ice cream, while $U(c_1)$ represent those immediate pleasures minus the future pain ξc_1 from the extra weight gained (see e.g. O'Donoghue and Rabin

2006 and Cremer and Pestieau 2011 for such an approach in the context of sin goods and savings, respectively). Then, $\boldsymbol{\tau}^b(\boldsymbol{q}, w) = \left(\tau_0^b, \tau_1^b\right)$ where $\tau_0^b = 0$ and $\tau_1^b = \xi > 0$, so that the misoptimization wedge τ_1^b on good 1 is exactly equal to the associated internality ξ.

Misperception model We turn to the misperception model, which was already outlined in Section 2. There are two primitives: a utility function $u(\boldsymbol{c})$ and a perception function $\boldsymbol{q}^s(\boldsymbol{q}, w)$. Given true prices $\boldsymbol{q}$, perceived prices $\boldsymbol{q}^s$, and budget w, the demand

$$\boldsymbol{c}^s(\boldsymbol{q}, \boldsymbol{q}^s, w) = \arg\max_{\boldsymbol{c}|\boldsymbol{q}^s} u(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q} \cdot \boldsymbol{c} \leq w$$

is the consumption vector $\boldsymbol{c}$ satisfying $u_{\boldsymbol{c}}(\boldsymbol{c}) = \lambda^s \boldsymbol{q}^s$ for some $\lambda^s > 0$ and $\boldsymbol{q} \cdot \boldsymbol{c} = \boldsymbol{w}$.[14] Then the primitive demand function $\boldsymbol{c}(\boldsymbol{q}, w)$ of the general model is given by

$$\boldsymbol{c}(\boldsymbol{q}, w) = \boldsymbol{c}^s(\boldsymbol{q}, \boldsymbol{q}^s(\boldsymbol{q}, w), w).$$

The misoptimization wedge is formally given by $\boldsymbol{\tau}^b(\boldsymbol{q}, w) = \boldsymbol{q} - \frac{\boldsymbol{q}^s(\boldsymbol{q},w)}{\boldsymbol{q}^s \cdot \boldsymbol{c}_w(\boldsymbol{q},w)}$. However, because in the misperception model, we have $\boldsymbol{q}^s(\boldsymbol{q}, w) \cdot \boldsymbol{S}^H(\boldsymbol{q}, w) = 0$, and because $\boldsymbol{\tau}^b(\boldsymbol{q}, w)$ enters all our formulas through $\boldsymbol{\tau}^b(\boldsymbol{q}, w) \cdot \boldsymbol{S}^H(\boldsymbol{q}, w)$, we can take

$$\boldsymbol{\tau}^b(\boldsymbol{q}, w) = \boldsymbol{q} - \boldsymbol{q}^s(\boldsymbol{q}, w) \tag{12}$$

in all of our formulas involving $\boldsymbol{\tau}^b \cdot \boldsymbol{S}^H$. This fleshes out the idea that $\boldsymbol{\tau}^b(\boldsymbol{q}, w)$ represents the impact of the discrepancy between true prices and perceived prices.

Given this demand function $\boldsymbol{c}(\boldsymbol{q}, w)$, the notions of the general model apply and receive a concrete interpretation. First, we define the matrix of marginal perceptions

$$M_{ij}(\boldsymbol{q}, w) = \frac{\partial q_i^s(\boldsymbol{q}, w)}{\partial q_j},$$

where the j-th column $\boldsymbol{M}_j(\boldsymbol{q}, w) = \frac{\partial \boldsymbol{q}^s(\boldsymbol{q},w)}{\partial q_j}$ encodes the marginal impact of a change in true price q_j on the perceived price $\boldsymbol{q}^s$.[15] It is convenient to define the Hicksian demand of a fictitious rational agent facing prices $\boldsymbol{q}^s$

$$\boldsymbol{h}^r(\boldsymbol{q}^s, \widehat{u}) = \arg\min_{\boldsymbol{c}} \boldsymbol{q}^s \cdot \boldsymbol{c} \text{ s.t. } u(\boldsymbol{c}) \geq \widehat{u},$$

and the associated Slutsky matrix

$$\boldsymbol{S}_j^r(\boldsymbol{q}^s, \widehat{u}) = \boldsymbol{h}_{q_j^s}^r(\boldsymbol{q}^s, \widehat{u}).$$

We also define $\boldsymbol{S}^r(\boldsymbol{q}, w) = \boldsymbol{S}^r(\boldsymbol{q}^s(\boldsymbol{q}, w), v(\boldsymbol{q}, w))$. Then, the utility-compensated Slutsky matrix

[14] If there are several such λ, we take the lowest one, which is also the utility-maximizing one.

[15] For instance, in specification (2), $M_{ij} = m_j 1_{\{i=j\}}$.

$\boldsymbol{S}^H(\boldsymbol{q}, w)$ of the general model is given by

$$\boldsymbol{S}^H(\boldsymbol{q}, w) = \boldsymbol{S}^r(\boldsymbol{q}, w) \cdot \boldsymbol{M}(\boldsymbol{q}, w). \tag{13}$$

i.e. $S^H_{ij}(\boldsymbol{q}, w) = \sum_k S^r_{ik}(\boldsymbol{q}, w) M_{kj}(\boldsymbol{q}, w)$. That is, when the price of good j changes, it creates a change $M_{kj}(\boldsymbol{q}, w) = \frac{\partial q^s_k(\boldsymbol{q}, w)}{\partial q_j}$ in the perceived prices q^s_k of a generic good k, which in turn changes (via the rational Slutstky matrix $\boldsymbol{S}^r(\boldsymbol{q}, w)$) the demand for good i.[16]

3.2 Optimal Taxation to Raise Revenues and Redistribute: Ramsey

There are H agents indexed by h. Each agent is competitive (price taker) as described in Section 3.1. All the functions describing the behavior and welfare of agents are allowed to depend on h.[17] We omit the dependence of all functions on $(\boldsymbol{q}, w)$, unless an ambiguity arises.

We introduce a social welfare function $W\left(v^1, ..., v^H\right)$ and a marginal value of public funds λ. The planning problem is

$$\max_{\boldsymbol{\tau}} L(\boldsymbol{\tau}),$$

where[18]

$$L(\boldsymbol{\tau}) = W\left(\left(v^h(\boldsymbol{p} + \boldsymbol{\tau}, w)\right)_{h=1...H}\right) + \lambda \sum_h \left[\boldsymbol{\tau} \cdot \boldsymbol{c}^h(\boldsymbol{p} + \boldsymbol{\tau}, w) - w\right].$$

Good 0 is constrained to be untaxed: $\tau_0 = 0$. We assume perfectly elastic supply with fixed producer prices $\boldsymbol{p}$. We relax this assumption in Section 7.1 where we consider the case of imperfectly elastic supply with endogenous producer prices $\boldsymbol{p}$.

Following Diamond (1975), we define γ^h to be the social marginal utility of income for agent h

$$\gamma^h = \beta^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_w, \tag{14}$$

where

$$\beta^h = W_{v^h} v^h_w \tag{15}$$

is the social marginal welfare weight. The difference $\lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_w$ between γ^h and β^h captures the marginal impact on tax revenues of a marginal increase in the income of agent h.

[16] There always exists a representation of the general model as a misperception model, but not as a decision vs. experienced utility model (see Lemma 12.1 in the online appendix).

[17] This formalism allows the interpretation that each agent has a type $t(h)$, that all agents of the same type are identical with the number of agents of type h given by H_h.

[18] The analysis is identical if we allow for endowments $\boldsymbol{e}^h$, using the objective function

$$L(\boldsymbol{\tau}) = W\left(\left(v^h\left(\boldsymbol{p} + \boldsymbol{\tau}, w + \boldsymbol{p} \cdot \boldsymbol{e}^h\right)\right)_{h=1...H}\right) + \lambda \sum_h \left[\boldsymbol{\tau} \cdot \boldsymbol{c}^h\left(\boldsymbol{p} + \boldsymbol{\tau}, w + \boldsymbol{p} \cdot \boldsymbol{e}^h\right) - w\right].$$

We also renormalize the misoptimization wedge

$$\widetilde{\boldsymbol{\tau}}^{b,h} = \frac{\beta^h}{\lambda}\boldsymbol{\tau}^{b,h}. \tag{16}$$

We now characterize the optimal tax system.[19]

Proposition 3.1 (Behavioral many-person Ramsey formula) *If commodity i can be taxed, then at the optimum*

$$\frac{\partial L\left(\boldsymbol{\tau}\right)}{\partial \tau_i} = 0 \quad \textit{with} \quad \frac{\partial L\left(\boldsymbol{\tau}\right)}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \lambda(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{S}_i^{C,h}]. \tag{17}$$

Proof We have

$$\frac{\partial L}{\partial \tau_i} = \sum_h [W_{v^h} v_w^h \frac{v_{q_i}^h}{v_w^h} + \lambda c_i^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}_{q_i}^h].$$

Using the definition of $\beta^h = W_{v^h} v_w^h$, the behavioral versions of Roy's identity (11), and the Slustky relation, we can rewrite this as

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

We then use the definition of the social marginal utility of income $\gamma^h = \beta^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}_w^h$ to get

$$\frac{\partial L}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \left[\lambda \boldsymbol{\tau} - \beta^h \boldsymbol{\tau}^{b,h}\right] \cdot \boldsymbol{S}_i^{C,h}].$$

The result follows using the renormalization (16) of the misoptimization wedge.

□

An intuition for this formula can be given along the following lines. The impact of a marginal increase in $d\tau_i$ on social welfare is the sum of three effects: a mechanical effect, a substitution effect, and a misoptimization effect.[20]

Let us start with the mechanical effect $\sum_h (\lambda - \gamma^h) c_i^h d\tau_i$. If there were no changes in behavior, then the government would collect additional revenues $c_i^h d\tau_i$ from agent h, which are valued by the

[19] Suppose that there is uncertainty, possibly heterogeneous beliefs, several dates for consumption, and complete markets. Then, our formula (17) applies without modifications, interpreting goods as a state-and-date contingent goods. See Spinnewijn (2014) for an analysis of unemployment insurance when agents misperceive the probability of finding a job, and Davila (2014) for an analysis of a Tobin tax in financial markets with heterogeneous beliefs.

[20] We can also write the optimal tax formula using the utility-compensated Slutsky matrices $\boldsymbol{S}^{H,h}$. For this purpose, it is convenient to introduce a different renormalization of the misoptimization wedge $\widetilde{\widetilde{\boldsymbol{\tau}}}^{b,h} = \frac{\gamma^h}{\lambda}\boldsymbol{\tau}^{b,h}$. We then get

$$0 = \frac{\partial L\left(\boldsymbol{\tau}\right)}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \lambda(\boldsymbol{\tau} - \widetilde{\widetilde{\boldsymbol{\tau}}}^{b,h}) \cdot \boldsymbol{S}_i^{H,h}]. \tag{18}$$

government as $(\lambda - \gamma^h)c_i^h d\tau_i$. Indeed, taxing one dollar from agent h to the government creates a net welfare change of $\lambda - \gamma^h$, where λ is the value of public funds and γ^h is the social marginal utility of income for agent h (which includes the associated income effect on tax revenues).

Let us turn to the substitution effect $\sum_h \lambda \boldsymbol{\tau} \cdot \boldsymbol{S}_i^{C,h} d\tau_i$. The change in consumer prices resulting from the tax change $d\tau_i$ induces a change in behavior $\boldsymbol{S}_i^{C,h} d\tau_i$ of agent h over and above the income effect accounted for in the mechanical effect. The resulting change $\boldsymbol{\tau} \cdot \boldsymbol{S}_i^{C,h} d\tau_i$ in tax revenues is valued by the government as $\lambda \boldsymbol{\tau} \cdot \boldsymbol{S}_i^{C,h} d\tau_i$.

Finally, let us analyze the misoptimization effect $-\sum_h \lambda \widetilde{\boldsymbol{\tau}}^{b,h} \cdot \boldsymbol{S}_i^{C,h} d\tau_i$. This effect is linked to the substitution effect. If agent h were rational, then the change in behavior captured by the substitution effect would have no first-order effects on his utility. This is a consequence of the envelope theorem. When agent h is behavioral, this logic fails, and the change in behavior associated with the substitution effect has first-order effects $-\beta^h \boldsymbol{\tau}^{b,h} \cdot \boldsymbol{S}_i^{C,h} d\tau_i = -\lambda \widetilde{\boldsymbol{\tau}}^{b,h} \cdot \boldsymbol{S}_i^{C,h} d\tau_i$ on his utility.

One way to think about the optimal tax formulas (17) is as a linear system of equations indexed by i in the optimal taxes τ_j for the different commodities

$$\frac{-\sum_h \boldsymbol{S}_{ji}^{C,h} \tau_j}{c_i} = 1 - \frac{\bar{\gamma}}{\lambda} - cov(\frac{\gamma^h}{\lambda}, \frac{H c_i^h}{c_i}) + \frac{-\sum_j \widetilde{\boldsymbol{\tau}}_j^{b,h} \boldsymbol{S}_{ji}^{C,h}}{c_i},$$

where $c_i = \sum_h c_i^h$ is total consumption of good i and $\bar{\gamma} = \frac{1}{H}\sum_h \gamma^h$ is the average social marginal utility of income. Of course the coefficients in this linear system of equations and the forcing terms (on the right-hand side) are endogenous and depend on taxes τ_j. Nevertheless, at the optimum, one can in principle solve out the linear system to express the taxes τ_j as a function of these coefficients and forcing terms (valued at optimal taxes). The first forcing term $1 - \frac{\bar{\gamma}}{\lambda} - cov(\frac{\gamma^h}{\lambda}, \frac{H c_i^h}{c_i})$ captures the revenue raising and redistributive objectives of taxation. The second forcing term $\frac{-\sum_j \tilde{\tau}_j^{b,h} \boldsymbol{S}_{ji}^{C,h}}{c_i}$ captures the corrective objective of taxation to address the effects of misoptimization. Note that the formulas feature $\widetilde{\boldsymbol{\tau}}^{b,h} = \frac{\beta^h}{\lambda} \boldsymbol{\tau}^{b,h}$ so that the misoptimization wedges are weighted by the social marginal welfare weights β^h. Intuitively, optimal taxes put more weight on addressing misoptimization by agents on which the social welfare function puts more weight.

Comparison with the traditional model with no behavioral biases In the traditional model of Diamond (1975) where all agents are rational, only the mechanical and substitution effects are present, yielding

$$0 = \frac{\partial L\left(\boldsymbol{\tau}\right)}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{S}_i^{r,h}].$$

Adding behavioral agents introduces the following differences. First, the changes in behavior (income and substitution effects) and the social marginal welfare weights are modified, leading to different values for β^h, γ^h, and a different Slutsky matrix $\boldsymbol{S}_i^{C,h}$. Second, there is a new effect (the misoptimization effect) leading to a new term $-\lambda \widetilde{\boldsymbol{\tau}}^{b,h} \cdot \boldsymbol{S}_i^{C,h}$.

Moreover, in the traditional model without behavioral biases we can use the symmetry of the Slutsky matrix $\boldsymbol{S}^{r,h}$ to write $\boldsymbol{\tau} \cdot \boldsymbol{S}_i^{r,h} = \sum_j \tau_j \boldsymbol{S}_{ji}^{r,h}$ as $\boldsymbol{\tau} \cdot \boldsymbol{S}_i^{r,h} = \sum_j \tau_j \boldsymbol{S}_{ij}^{r,h}$. We can then rewrite the optimal tax formula in "discouragement" form as

$$\frac{-\sum_j \tau_j \boldsymbol{S}_{ij}^{r,h}}{c_i} = 1 - \frac{\bar{\gamma}}{\lambda} - cov(\frac{\gamma^h}{\lambda}, \frac{Hc_i^h}{c_i}),$$

The left-hand side is the discouragement index of good i, which loosely captures how much the consumption of good i is discouraged by the taxes τ_j on all the different commodities j. The right-hand side indicates that in the absence of distributive concerns (homogenous $\gamma^h = \gamma$), all goods should be uniformly discouraged in proportion to the relative intensity $1 - \frac{\bar{\gamma}}{\lambda}$ of the raising revenue objective. With redistributive concerns (heterogenous γ^h), goods that are disproportionately consumed by agents that society tries to redistribute towards (agents with a high γ^h) should be discouraged less.

With behavioral biases, the optimal tax formula cannot in general be simply written in discouragement form. This is because the Slutsky matrix $\boldsymbol{S}^{C,h}$ is not symmetric in general. But when the the Slutsky matrix $\boldsymbol{S}^{C,h}$ is symmetric, we can go through the same steps as above and rewrite the optimal tax formula (17) as follows

$$\frac{-\sum_j \tau_j \boldsymbol{S}_{ij}^{C,h}}{c_i} = 1 - \frac{\bar{\gamma}}{\lambda} - cov\left(\frac{\gamma^h}{\lambda}, \frac{Hc_i^h}{c_i}\right) - \frac{\sum_j \tilde{\tau}_j^{b,h} \boldsymbol{S}_{ij}^{C,h}}{c_i},$$

with a similar interpretation. The difference is once again that the values for β^h, γ^h, and for the Slutsky matrix $\boldsymbol{S}_{ji}^{C,h}$ are different than in the traditional model with no behavioral biases, and that there are new terms $\tilde{\tau}_j^{b,h}$ reflecting misoptimization.

Lump-sum taxes and the negative income tax Suppose that in addition to linear commodity taxes, the government can use a lump-sum tax or rebate, identical for all agents (a negative income tax). This amounts to assuming that the government can adjust w. Then optimal commodity taxes are characterized by the exact same conditions. But there is now an additional optimality condition corresponding to the optimal choice of the lump-sum rebate w yielding

$$\bar{\gamma} = \frac{1}{H} \sum_{h=1}^{H} \gamma^h = \lambda. \tag{19}$$

Suppose that agents are homogeneous, and that lump-sum taxes are available, so that $\lambda = \gamma^h$. Proposition 3.1 implies that the optimal tax satisfies $\boldsymbol{\tau} = \tilde{\boldsymbol{\tau}}^{b,h}$. The optimal tax corrects the agent's internality. If in additional all agents are rational, then we get $\boldsymbol{\tau} = \tilde{\boldsymbol{\tau}}^{b,h} = 0$.

When agents are heterogenous, formula (16) implies that optimal taxes are not zero anymore $\boldsymbol{\tau} \neq 0$. The optimal tax formula then captures a version of the celebrated negative income tax famously proposed by Milton Friedman.

3.3 Optimal Taxation with Externalities: Pigou

We now introduce externalities and study the consequences for the optimal design of commodity taxes. The utility of agent h is now $u^h\left(\boldsymbol{c}^h, \xi\right)$, where $\xi = \xi\left(\left(\boldsymbol{c}^h\right)_{h=1...H}\right)$ is a one-dimensional externality (for simplicity) that depends on the consumption vectors of all agents and is therefore endogenous to the tax system.[21] All individual functions encoding the behavior and welfare of agents now depend on the externality ξ.

The planning problem becomes

$$\max_{\boldsymbol{\tau}} L\left(\boldsymbol{\tau}\right),$$

where

$$L\left(\boldsymbol{\tau}\right) = W\left(\left(v^h\left(\boldsymbol{p}+\boldsymbol{\tau}, w, \xi\right)\right)_{h=1...H}\right) + \lambda \sum_h \left[\boldsymbol{\tau}\cdot \boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau}, w, \xi\right) - w\right],$$

$$\xi = \xi\left(\left(\boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau}, w, \xi\right)\right)_{h=1...H}\right).$$

We define the social marginal value of the externality

$$\Xi = \frac{\sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda \boldsymbol{\tau}\cdot \boldsymbol{c}_\xi^h\right]}{1 - \sum_h \xi_{\boldsymbol{c}^h} \boldsymbol{c}_\xi^h}.$$

This definition includes all the indirect effects of the externality on consumption and the associated effects on tax revenues (the term $\lambda\boldsymbol{\tau}\cdot\boldsymbol{c}_\xi^h$ in the numerator) as the associated multiple round effects on the externality (the "multiplier" term encapsulated in the denominator). With this convention, Ξ is negative for a bad externality, like pollution. We also define the (agent-specific) Pigouvian wedge

$$\boldsymbol{\tau}^{\xi,h} = -\frac{\Xi \xi_{\boldsymbol{c}^h}}{\lambda}.$$

It represents the dollar value of the externality created by one more unit of consumption by agent h. We finally define the externality-augmented social marginal utility of income

$$\gamma^{\xi,h} = \gamma^h + \Xi \xi_{\boldsymbol{c}^h} \boldsymbol{c}_w^h = \beta^h + \lambda\left(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h}\right)\cdot \boldsymbol{c}_w^h.$$

This definition captures the fact that, as one dollar is given to the agent, his direct social utility increases by γ^h, but the extra dollar changes consumption by $\boldsymbol{c}_w^h$, and, hence, the total externality by $\xi_{\boldsymbol{c}^h}\boldsymbol{c}_w^h$, with a welfare impact $\Xi\xi_{\boldsymbol{c}^h}\boldsymbol{c}_w^h$. The next proposition generalizes Proposition 3.1.[22]

[21] For example, to capture an externality (e.g. second hand smoke) from the consumption of good 1, we could specify $\xi = \frac{\xi_*}{H}\sum_h c_1^h$ and $u^h\left(\boldsymbol{c}^h, \xi\right) = u^h\left(\boldsymbol{c}^h\right) - \xi$.

[22] Formally, misoptimization and externality wedges ($\tilde{\boldsymbol{\tau}}^{b,h}$, $\boldsymbol{\tau}^{\xi,h}$) enter symmetrically in the optimal tax formula. In some particular cases, behavioral biases can be alternatively modelled as externalities (for example, this is the case for a decision vs. experienced utility model with a representative agent). But this is not true in general. For example, misperceptions naturally give rise to non-symmetric Slutsky matrices $\boldsymbol{S}_i^{C,h}$ which cannot be captured with a traditional externality model. Moreover, even with a quasilinear utility function and separable utility (so that the Slutisky matrix is diagonal and hence symmetric), the misperception model would require externalities that directly

Proposition 3.2 (Behavioral many-person Pigou formula) *If commodity i can be taxed, then at the optimum*

$$\frac{\partial L(\boldsymbol{\tau})}{\partial \tau_i} = 0, \quad with \quad \frac{\partial L(\boldsymbol{\tau})}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^{\xi,h}\right) c_i^h + \lambda \left(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h} - \boldsymbol{\tau}^{\xi,h}\right) \cdot \boldsymbol{S}_i^{C,h}]. \tag{20}$$

3.4 Optimal Nudges

In this section, we develop a model of nudges (Thaler and Sunstein 2008) and we derive a formula characterizing optimal nudges. At an abstract level, we assume that a nudge influences consumption but does not enter the budget constraint. This is the key difference between a nudge and a tax. The demand function $\boldsymbol{c}^h(\boldsymbol{q}, w, \chi)$ satisfies the budget constraint $\boldsymbol{q} \cdot \boldsymbol{c}^h(\boldsymbol{q}, w, \chi) = w$, where χ is the nudge vector.

Specializing the general model a bit more, we propose the following concrete model of nudges. In the absence of a nudge, the agent has decision utility $u^{s,h}$ and perceived price $\boldsymbol{q}^{s,h,*}$, which is the perceived price before the nudge. With the nudge,

$$\boldsymbol{c}^h(\boldsymbol{q}, w, \chi) = \arg\operatorname*{smax}_{\boldsymbol{c} | B^{s,h}} u^{s,h}(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q} \cdot \boldsymbol{c} \leq w.$$

That is, his demand $\boldsymbol{c}$ satisfies $u_{\boldsymbol{c}}^{s,h}(\boldsymbol{c}) = \Lambda^h B_{\boldsymbol{c}}^{s,h}(\boldsymbol{q}, \boldsymbol{c}, \chi)$ for a constant $\Lambda^h > 0$ pinned down by the budget constraint $\boldsymbol{q} \cdot \boldsymbol{c} = w$. The perceived (potentially nonlinear) budget constraint $B^{s,h}(\boldsymbol{q}, \boldsymbol{c}, \chi) \leq w$ is modified by the nudge. But the true budget constraint is unchanged.

A useful specification is that of a nudge as a (psychological) tax on consumption good i, in which case

$$B^{s,h}(\boldsymbol{q}, \boldsymbol{c}, \chi) = \boldsymbol{q}^{s,h,*} \cdot \boldsymbol{c} + \chi \eta^h c_i, \tag{21}$$

where $\eta^h \geq 0$ captures the nudgeability of the agent ($\eta^h = 0$ corresponding to a non-nudgeable agent).[23] An example of such nudge is a public campaign against cigarettes ($\chi > 0$) or for recycling ($\chi < 0$). Another useful specification is that of a nudge as a (psychological) anchor on a given consumption of good i, in which case

$$B^{s,h}(\boldsymbol{q}, \boldsymbol{c}, \chi) = \boldsymbol{q}^{s,h,*} \cdot \boldsymbol{c} + \eta^h |c_i - \chi|, \tag{22}$$

so that an extra psychological penalty if the agent deviates from the quantity χ recommended by the nudge. For instance, the nudge could be a default allocation in a retirement plan (see e.g. Carroll et al. 2009). With the second specification, a number of agents will choose the default.[24]

depend on price wedge $\boldsymbol{q} - \boldsymbol{q}^s$, which is not covered in the traditional externalities literature.

[23] We could generalize to an costly attention away from the nudge along the lines of Section 6, e.g. with $\eta^h(m_\chi)$ weakly decreasing in m_χ, with $m_\chi = 1$ being an non-nudgeable agent.

[24] The technical condition for the default to be chosen is $\left| u_{c_i}^h(\chi_i)/\Lambda^h - q_i^{s,h,*} \right| < \eta^h$.

A nudge may also directly affect agents' utility $u^h(\boldsymbol{c},\chi)$. This could be captured by specifying $u^h(\boldsymbol{c},\chi) = u(\boldsymbol{c}) - \iota^h \chi c_i$ or perhaps $u^h(\boldsymbol{c},\chi) = u(\boldsymbol{c}) - \iota^h |c_i - \chi|$, depending on the context. For example, a display of cancerous lungs on a pack of cigarettes not only nudges people away from consuming a cigarette (as captured by η^h), but may also directly lower their utility, as measured by ι^h.[25]

We now characterize optimal nudges.

Proposition 3.3 (Optimal nudge formula) *Optimal nudges satisfy*

$$\frac{\partial L(\boldsymbol{\tau},\chi)}{\partial \chi} = 0, \quad \textit{with} \quad \frac{\partial L}{\partial \chi}(\boldsymbol{\tau},\chi) = \sum_h [\lambda(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{c}^h_\chi + \beta^h \frac{u^h_\chi}{v^h_w}]. \tag{23}$$

The optimality conditions for taxes $\frac{\partial L(\boldsymbol{\tau},\chi)}{\partial \tau_i} = 0$ *are unchanged.*

Proof We use the fact that $\boldsymbol{q} \cdot \boldsymbol{c}(\boldsymbol{q}, w, \chi) = w$ implies $\boldsymbol{q} \cdot \boldsymbol{c}_\chi = 0$:

$$\begin{aligned}\frac{\partial L}{\partial \chi} &= \sum_h [W_{v^h}\left(v^h_w \frac{u^h_{\boldsymbol{c}}}{v^h_w} + v^h_w \frac{u^h_\chi}{v^h_w}\right) + \lambda(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi h})]\boldsymbol{c}^h_\chi = \sum_h [\beta^h \left(\frac{u^h_{\boldsymbol{c}}}{v^h_w} - \boldsymbol{q} + \boldsymbol{q} + \frac{u^h_\chi}{v^h_w}\right) + \lambda(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi h})]\boldsymbol{c}^h_\chi \\ &= \sum_h \left[[\beta^h \left(\frac{u^h_{\boldsymbol{c}}}{v^h_w} - \boldsymbol{q}\right) + \lambda \boldsymbol{\tau}]\boldsymbol{c}^h_\chi + \beta^h \frac{u^h_\chi}{v^h_w}\right] = \sum_h \left[[-\lambda\left(\boldsymbol{\tau}^{\xi,h} + \widetilde{\boldsymbol{\tau}}^{b,h}\right) + \lambda \boldsymbol{\tau}]\boldsymbol{c}^h_\chi + \beta^h \frac{u^h_\chi}{v^h_w}\right].\end{aligned}$$

□

This formula clarifies how to best use nudges. It has four terms corresponding to three potentially conflicting goals of nudges. The first term, $\lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_\chi$, captures the fact that the changes in behavior induced by nudges directly change tax revenues. The second term, $-\lambda \boldsymbol{\tau}^{\xi,h} \cdot \boldsymbol{c}^h_\chi$, captures the fact that the changes in behavior induced by nudges affect welfare and tax revenues through their effect on externalities. The third term, $-\lambda \widetilde{\boldsymbol{\tau}}^{b,h} \cdot \boldsymbol{c}^h_\chi$, captures the fact that the changes in behavior induced by nudges affect welfare because agents misoptimize. The fourth term, $\beta^h \frac{u^h_\chi}{v^h_w}$, captures the potential direct effects of nudges on utility.[26]

[25] Glaeser (2006) and Loewenstein and O'Donoghue (2006) discuss the notion that nudges have a psychic cost.

[26] At this level of generality, quantity restrictions (e.g. mandates or caps) could be modeled in as nudges, with the vector χ representing a vector of quantity restrictions influencing the demand functions $\boldsymbol{c}^h(\boldsymbol{q}, w, \chi)$. Of course the particular way in which quantity restrictions influence demand differs from nudges, and as a result the implications of these first order conditions are very different for nudges and quantity restrictions. For example, imagine that χ indexes a quantity restriction stipulating that the consumption of a certain commodity i must be equal to χ. In the context of the specialized model outlined above, this would lead to $\boldsymbol{c}^h(\boldsymbol{q}, w, \chi) = \arg\operatorname{smax}_{\boldsymbol{c}|\boldsymbol{q}^{s,h,*}} u^{s,h}(\boldsymbol{c})$ s.t. $\boldsymbol{q} \cdot \boldsymbol{c} \leq w$ and $c_i = \chi$. That is, his demand $\boldsymbol{c}$ satisfies $c_i = u^{s,h}_{c_j}(\boldsymbol{c})$ and $c_j = \Lambda^h_j q^{s,h,*}$ for all $j \neq i$, for a constant $\Lambda^h > 0$ pinned down by the budget constraint $\boldsymbol{q} \cdot \boldsymbol{c} = w$.

3.5 A Useful Simple Case

In this section, we work out a useful particularization of the general model which yields simple optimal tax formulas. This simple case will prove useful to construct many of our examples in Section 4. It also allows to explicitly link Sections 2 and 3 by showing how to obtain the tax formulas of Section 2 as special cases of the general tax formulas derived in Section 3 in the limit of small taxes.

We use a hybrid model with both decision vs. experienced utility and misperceptions. We make two simplifying assumptions. First, we assume that decision vs. experienced utility are quasilinear so that the marginal utility of wealth is constant. Second, we assume that misperceptions are constant. Finally, we allow for externalities ξ but assume that they are separable from consumption.

Formally, we decompose consumption $\boldsymbol{c}=(c_0,\boldsymbol{C})$ where $\boldsymbol{C}$ is of dimension n, and we normalize $p_0=q_0=1$. The experienced utility of agent h is quasilinear

$$u^h(\boldsymbol{c}_0,\boldsymbol{C},\xi)=c_0+U^h(\boldsymbol{C})-\xi,$$

where $\xi=\xi(\left(\boldsymbol{C}^h\right)_{h=1...H})$. Agent h is subject to two sets of biases. First, he maximizes a decision utility

$$u^{s,h}(\boldsymbol{c}_0,\boldsymbol{C},\boldsymbol{\xi})=c_0+U^{s,h}(\boldsymbol{C})-\xi$$

which differs from his experienced utility, but remains quasilinear.[27] Second, he perceives prices to be $\boldsymbol{q}^{s,h}=\boldsymbol{p}+\boldsymbol{\tau}^{s,h}$ with $\boldsymbol{\tau}^{s,h}=\boldsymbol{M}^h\boldsymbol{\tau}$, so that the misperception of taxes is given by a constant matrix of marginal perceptions $\boldsymbol{M}^h$. The corresponding perception function is[28]

$$\boldsymbol{q}^{s,h}(\boldsymbol{q})=\boldsymbol{p}+\boldsymbol{M}^h(\boldsymbol{q}-\boldsymbol{p}).$$

The demand $\boldsymbol{c}^h(\boldsymbol{q},w,\xi)=\left(c_0^h(\boldsymbol{q},w),\boldsymbol{C}^h(\boldsymbol{q})\right)$ of agent h is such that $\boldsymbol{C}^h(\boldsymbol{q})=\boldsymbol{C}^{s,h}\left(\boldsymbol{q}^{s,h}(\boldsymbol{q})\right)$ and $c_0^h(\boldsymbol{q},w)=w-\boldsymbol{q}\cdot\boldsymbol{C}^h(\boldsymbol{q})$, where $\boldsymbol{C}^{s,h}\left(\boldsymbol{q}^{s,h}\right)=\arg\max_{\boldsymbol{C}}U^{s,h}(\boldsymbol{C})-\boldsymbol{q}^{s,h}\cdot\boldsymbol{C}$.

We define the rational Slutsky matrix

$$\boldsymbol{S}^{r,h}\left(\boldsymbol{q}^{s,h}\right)=\frac{\partial\boldsymbol{C}^{s,h}\left(\boldsymbol{q}^{s,h}\right)}{\partial\boldsymbol{q}^{s,h}}.$$

This corresponds to the Slutsky matrix of an agent with decision utility $U^{s,h}$, but who perceives taxes correctly. Because decision utility is quasilinear, there are no income effects and we have

$$\boldsymbol{S}^{H,h}(\boldsymbol{q},w)=\boldsymbol{S}^{C,h}(\boldsymbol{q},w)=\boldsymbol{S}^{r,h}\left(\boldsymbol{q}^{s,h}(\boldsymbol{q})\right)\cdot\boldsymbol{M}^h.$$

[27] When choosing his consumption, the agent does not internalize the effect of his decisions on the value of ξ.

[28] In all those definitions, we omit the row and columns corresponding to good 0, which has no taxes and no misperceptions.

We also define the internality wedge $\boldsymbol{\tau}^{I,h}$ and the internality/externality wedge $\boldsymbol{\tau}^{X,h}$ as follows:

$$\boldsymbol{\tau}^{I,h} = U_{\boldsymbol{C}}^{s,h}(\boldsymbol{C}) - U_{\boldsymbol{C}}^{h}(\boldsymbol{C}) \quad \text{and} \quad \boldsymbol{\tau}^{X,h} = \frac{\beta^h}{\lambda}\boldsymbol{\tau}^{I,h} + \boldsymbol{\tau}^{\xi,h}.$$

The wedge $\boldsymbol{\tau}^{I,h}$ is closely related to the misoptimization wedge $\boldsymbol{\tau}^{b,h}$ according to

$$\boldsymbol{\tau}^{b,h} = \boldsymbol{\tau}^{I,h} + \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h}.$$

Basically, $\boldsymbol{\tau}^{b,h}$ captures two forms of misoptimization: those arising from the difference between decision and experienced utility ($\boldsymbol{\tau}^{I,h}$) and those arising from the misperception of taxes ($\boldsymbol{\tau} - \boldsymbol{\tau}^{s,h}$). In this example, we find it useful to separate them.

We now characterize optimal taxes. Because there are no wealth effects in consumption, we have $\gamma^{\xi,h} = \gamma^h = \beta^h$.

Proposition 3.4 (Optimal tax formula with constant marginal utility of wealth and constant misperceptions) *In the constant marginal utility of wealth and constant misperceptions specification of the general model, optimal taxes satisfy*

$$\boldsymbol{\tau} = -(\sum_h \boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}(I - \left(I - \boldsymbol{M}^h\right)\frac{\gamma^h}{\lambda}))^{-1}\sum_h[(1-\frac{\gamma^h}{\lambda})\boldsymbol{C}^h + \boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}\boldsymbol{\tau}^{X,h}] \tag{24}$$

This formula is a direct application of the tax formulas in Propositions 3.1 and 3.2, obtained by particularizing the general model, and by solving the system of linear equations in taxes $\boldsymbol{\tau}$ formed by these tax formulas.

In the context of this simple case, we find it convenient to treat differently the different sources of misoptimization wedges $\boldsymbol{\tau}^{b,h}$. The internality wedges $\boldsymbol{\tau}^{I,h}$ are incorporated into the internality/externality wedges $\boldsymbol{\tau}^{X,h}$, while the misperception wedges $\boldsymbol{\tau} - \boldsymbol{\tau}^{s,h}$ are accounted for by the misperception matrices $\boldsymbol{M}^h$.

The tax formula (24) simplifies in the limit of small taxes, with small redistributive, revenue raising, and internality/externality correction motives. We have

$$\boldsymbol{\tau} = -(\sum_h \boldsymbol{S}^{\boldsymbol{M},r,h})^{-1}\sum_h[(1-\frac{\gamma^h}{\lambda})\boldsymbol{C}^h + \boldsymbol{S}^{\boldsymbol{M},r,h}\boldsymbol{\tau}^{X,\boldsymbol{M},h}] + O\left(\varepsilon^2\right), \tag{25}$$

where $\boldsymbol{S}^{\boldsymbol{M},r,h} = \boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}\boldsymbol{M}^h$ and $\boldsymbol{\tau}^{X,\boldsymbol{M},h} = \left(\boldsymbol{M}^h\right)^{-1}\boldsymbol{\tau}^{X,h}$ are the misperception-adjusted Slutsky matrix and internality/externality wedge, and $\varepsilon = \max\{\max_h\left|\frac{\gamma^h}{\lambda} - 1\right|, \max_h\left\|\boldsymbol{\tau}^{X,\boldsymbol{M},h}\right\|\}$ captures the strength of the redistribution, revenue raising, and internalities/externalities correction motives for taxation.[29]

[29] This formula coincides with the standard tax formula in the absence of behavioral biases for some fictitious rational agents with Slutsky matrices $\boldsymbol{S}^{\boldsymbol{M},r,h}$ and ideal externality wedges $\boldsymbol{\tau}^{X,\boldsymbol{M},h}$. However, the analogy with fictitious agents is misleading. Indeed, and importantly, the sensitivities of consumption to tax changes are given by

We call this the limit of small taxes for short. The tax formulas of Section 2 are special cases of equation (25). For example, the basic Ramsey tax formula in Proposition 2.1 obtains when there is no heterogeneity (all agents are identical), the Slutsky matrices are diagonal $\boldsymbol{S}^{r,h} = diag(y_1\psi_1, y_2\psi_2, ..., y_n\psi_n)$ where $y_i = \boldsymbol{C}_i^{s,h}(\boldsymbol{p})$, the misperception matrices are diagonal $\boldsymbol{M}^h = diag(m_1, m_2, ..., m_n)$, there are no internalities/externalities $\boldsymbol{\tau}^{X,h} = 0$, and with $\Lambda = \frac{\lambda - \gamma^h}{\gamma^h}$. Similarly, the basic Pigou formula in Proposition 2.2 obtains when in addition there is only one taxed good $n = 1$, there is no revenue raising motive for taxation $\lambda = \gamma^h$, and there is a constant internality/externality wedge $\tau^{X,h} = \xi$. Finally, the basic Pigou formula allowing for heterogeneity in attention or externality/internality in Proposition 2.3 obtains when there is heterogeneity across agents in $\boldsymbol{M}^h = m_1^h$ and in $\tau^{X,h} = \xi^h$.

4 Examples

In this section, we analyze different specializiations of the general model in order to extract concrete insights from the optimal tax formulations of the previous section.

4.1 Correcting Internalities/Externalities: Relaxation of the Principle of Targeting

The classical "principle of targeting" can be stated as follows. If the consumption of a good entails an externality, the optimal policy is to tax it, and not to subsidize substitute goods or tax complement goods. For example, if fuel pollutes, then optimal policy requires taxing fuel but not taxing fuel inefficient cars or subsidizing solar panels (see Salanié (2011) for such an example). Likewise, if fatty foods are bad for consumers, and they suffer from an internality, then fatty foods should be taxed, but lean foods should not be subsidized. As we shall see, misperceptions of taxes lead to a reconsideration of this principle of targeting.

We use the specialization of the general model developed in Section 3.5. We assume that $\gamma^h = \lambda$, so there is no revenue-raising motive and no redistribution motive. We also assume that the internality/externality wedge is constant across agents $\boldsymbol{\tau}^{X,h} = \boldsymbol{\tau}^X$. Equation (24) then yields the optimal tax:

$$\boldsymbol{\tau} = \left(\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^r\boldsymbol{M}^h\right]\right)^{-1}\mathbb{E}\left[\boldsymbol{M}^{h\prime}\right]\boldsymbol{S}^r\boldsymbol{\tau}^X.$$

We consider the case with $n = 2$ taxed goods (in addition to the untaxed good 0), where the consumption of good 1 features an internality/externality so that $\boldsymbol{\tau}^X = (\xi_*, 0)$ with $\xi_* > 0$. This can be generated as follows in the model in in Section 3.5. In the externality case, we simply assume that $\xi(\left(\boldsymbol{C}^h\right)_{h=1...H}) = \xi_* \frac{1}{H}\sum_h C_1^h$. In the internality case, we assume that $U^h(\boldsymbol{C}) = U^{s,h}(\boldsymbol{C}) - \xi_* C_1^h$. For example, in the externality case, good 1 could be fuel and good 2 a solar panel. In the internality

$\boldsymbol{S}^{C,h} = \boldsymbol{S}^{H,h} = \boldsymbol{S}^{r,h}\boldsymbol{M}^h$ and not by $\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}\boldsymbol{M}^h$.

example, good 1 could be fatty beef and good 2 lean turkey. In addition, we assume that the attention matrices are diagonal so that $\boldsymbol{M}^h = diag(m_1^h, m_2^h)$.

When agents have uniform misperceptions ($\boldsymbol{M}^h = \boldsymbol{M}$), the optimal tax is $\boldsymbol{\tau} = \boldsymbol{M}^{-1}\boldsymbol{\tau}^X$. 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.

Proposition 4.1 (Modified principle of targeting) *Suppose that the consumption of good* 1 *(but not good* 2*) entails a negative internality/externality. If agents perceive taxes correctly, then good* 1 *should be taxed, but good* 2 *should be left untaxed—the classical principle of targeting holds. If agents' misperceptions of the tax on good* 1 *are heterogeneous (var* $\left(m_1^h\right) > 0$*), and if the misperceptions* m_1^h *and* m_2^h *of the two goods are not too correlated, then, good* 2 *should be subsidized (respectively taxed) if and only if goods* 1 *and* 2 *are substitutes (respectively complements).*[30]

Proposition 4.1 shows that if people have heterogeneous attention to a fuel tax, then solar panels should be subsidize (this result is also reminiscent of Allcott, Mullainathan and Taubinsky (2014)).[31] The reason is that the tax on good 1 is an imperfect instrument in the presence of attention heterogeneity. It should therefore be supplemented with a subsidy on substitute goods and a tax on complement goods. A fuel tax should therefore be supplemented with a subsidy on solar panels and tax on fuel inefficient cars. Similarly, a fat tax should be supplemented withy a subsidy on lean foods.

A similar logic applies in the traditional model with no behavioral biases, if there is an externality, and this externality is heterogenous across agents. Our result should therefore be interpreted as an additional and potentially important reason why the principle of targeting might fail in the presence of behavioral biases: heterogenous perceptions of corrective taxes.

4.2 Internalities and Redistribution

Suppose that the poor consume "too many" sugary sodas. This brings up a difficult trade-off. On the one hand, taxing sugary sodas corrects the poor's internality. On the other hand, taxing sugary sodas redistributes away from the poor.[32]

To gain insights on how to balance these two conflicting objectives, we use the specialization of the general model developed in Section 3.5. For simplicity, we assume that good 1 is solely consumed by a class of agents, h^* but not by other agents $h \neq h^*$. We also assume that good 1 is separable, $U^{s,h^*}(\boldsymbol{C}) = U_1^{s,h^*}(c_1) + U_2^{s,h^*}(\boldsymbol{C}_2)$, where $\boldsymbol{C}_2 = (c_i)_{i\geq 2}$ and $U^{s,h}(\boldsymbol{C}) = U_2^{s,h}(\boldsymbol{C}_2)$ for $h \neq h^*$. We assume that good 1 features a harmful internality $\tau_1^{X,h^*} = \frac{\beta^{h^*}}{\lambda}\tau_1^{I,h^*} > 0$ and

[30] The required formal condition on the correlations of the misperceptions m_1^h and m_2^h of the two goods is $\mathbb{E}\left[m_2^h - \frac{\mathbb{E}\left[m_1^h m_2^h\right]}{\mathbb{E}\left[\left(m_1^h\right)^2\right]} m_1^h\right] > 0$.

[31] Recall that goods 1 and 2 are substitutes (respectively complements) if and only if $S_{12}^r > 0$ (respectively $S_{12}^r < 0$).

[32] This was actually the debate about a recent proposal in New York City.

that all taxes are correctly perceived. Formally, we take $U^h(\boldsymbol{C}) = U^{s,h}(\boldsymbol{C})$ for $h \neq h^*$ and $U^{h^*}(\boldsymbol{C}) = U^{s,h^*}(\boldsymbol{C}) - \tau_1^{I,h^*} c_1$. We denote by $\psi_1^{h^*} = -\frac{q_1}{c_1} S_{11}^{C,h^*}$ the elasticity of the demand for good 1 by agent h^*. As a concrete example, h^* could stand for "poor" and good 1 for "sugary sodas". Proposition 3.4 yields the following.

Proposition 4.2 (Taxation with both redistributive and corrective motives) *Suppose that good* 1 *is consumed only by agent* h^**, and entails an internality (captured by the internality wedge* τ_1^{I,h^*}*). Then the optimal tax on good 1 is*

$$\frac{\tau_1}{q_1} = (1 - \frac{\gamma^{h^*}}{\lambda}) \frac{1}{\psi_1^{h^*}} + \frac{\gamma^{h^*}}{\lambda} \frac{\tau_1^{I,h^*}}{q_1}. \tag{26}$$

The sign of the tax τ_1 is ambiguous because there are two forces at work, corresponding to the two terms on the right-hand side.[33] The first term $(1 - \frac{\gamma^{h^*}}{\lambda}) \frac{1}{\psi_1^{h^*}}$ corresponds to the redistributive objective of taxes. It is decreasing in the marginal social welfare weight γ^{h^*} on agent h^* and negative if $\gamma^{h^*} > \lambda$. This is because good 1 is consumed only by agent h^* and therefore taxing good 1 redistributes away from agent h^*. In absolute value, it is decreasing in the elasticity $\psi_1^{h^*}$ of the demand of good 1 by agent h^* because taxing or subsidizing the consumption of good 1 entails larger distortions when his consumption is more elastic.

The second term $\frac{\gamma^{h^*}}{\lambda} \frac{\tau_1^{I,h^*}}{q_1}$ corresponds to the internality-corrective motive of taxes. It is positive and increasing in γ^{h^*}. This is because good 1 is harmful as it entails a negative internality.

As we increase the weight γ^{h^*} on agent h^*, the tax τ_1 on good 1 increases if and only if $\frac{\tau_1^{I,h^*}}{q_1} > \frac{1}{\psi_1^{h^*}}$, i.e. if and only if the internality wedges is large enough and the demand elasticity is large enough.

This situation highlights one of the advantages of nudges over corrective taxes: they allow the correction of internalities while avoiding the associated mechanical redistributive effects of correctives taxes. To formalize this idea, imagine that the internality can be completely eliminated by an informational nudge (e.g. explaining the bad consequences of sugar), resulting in $\tau_1^{I,h^*} = 0$. Optimal taxes are then only used for redistribution, $\tau_1 = \frac{\lambda - \gamma^{h^*}}{\lambda \psi_1^{h^*}}$. In some cases, quantity restrictions can be used to the same effect, and these considerations may also help rationalize why governments choose to ban high-interest rate loans, rather than tax them, since taxes create an additional burden for the poor.

4.3 Nudges and Taxes

In this section, we propose an example illustrating the optimal determination of nudges and their interaction with taxes. We investigate how much optimal policy relies on taxes versus nudges. We examine whether nudges are complements or substitutes. And we illustrate some of the determinants of the choice between nudges and taxes.

[33] In independent work, Lockwood and Taubinsky (2015b) examine a related problem, in the context of a Mirrleesian income tax.

Optimal nudges We use the specialization of the general model developed in Section 3.5. There is only one taxed good $n = 1$.

We use quadratic utilities. Specifically, in the internality case, we assume that $U^{s,h}(c) = \frac{a^h c - \frac{1}{2}c^2}{\Psi}$ and $U^h(c) = \frac{a^h c - \frac{1}{2}c^2}{\Psi} - \tau^{I,h} c$. In the externality case, we assume that $U^h(c) = U^{s,h}(c) = \frac{a^h c - \frac{1}{2}c^2}{\Psi} - \xi$, where $\xi = \frac{\lambda}{\sum_h \beta^h} \sum_h \boldsymbol{\tau}^{\xi,h} c^h$. Recall that we define $\tau^{X,h} = \frac{\beta^h}{\lambda}\tau^{I,h} + \tau^{\xi,h}$.

We model the nudge as a psychological tax, as in (21). The demand of a consumer can then be expressed as

$$c^h(\tau, \chi) = c_0^h - \Psi\left(\eta^h \chi + m^h \tau\right),$$

where η^h is the sensitivity of agent h and m^h is the attention to the tax.

We start with the case where the nudge is the only instrument. We apply the optimal nudge formula (23).

Proposition 4.3 *When the nudge is the only instrument, the optimal nudge is given by*[34]

$$\chi = \frac{\mathbb{E}\left[\tau^{X,h}\eta^h\right]}{\mathbb{E}\left[{\eta^h}^2\right]}. \tag{27}$$

where $\mathbb{E}$ *denotes the average over agents* h*.*

If all agents have the same sensitivity $\eta_h = \eta$, the nudge is $\chi = \frac{\mathbb{E}\left[\tau^{X,h}\right]}{\eta}$, so the nudge is greater when the average internality/externality wedge is greater, and when agents are less nudgeable. Heterogeneities in nudgeability determine how well targeted the nudge is to the internality/externality. Controlling for $\mathbb{E}\left[\tau^{X,h}\eta^h\right]$, the nudge is weaker when there is more heterogeneity in nudgeability (higher $\mathbb{E}\left[{\eta^h}^2\right]$), and, controlling for heterogeneity in nudgeability, the nudge is stronger when nudgeability is correlated with the internality/externality ($\mathbb{E}\left[\tau^{X,h}\eta^h\right]$ is higher).

Jointly optimal nudges and taxes We next consider the optimal joint policy using both nudges and taxes. We only highlight a few results; more results can be found in the online appendix (Section 12.4). One can show that

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

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 $\mathbb{E}\left[\left(\lambda - \gamma^h\left(1 - m^h\right)\right)\eta^h\right] \leq 0$. 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

[34] The intermediate steps are as follows. Using $\boldsymbol{c}_\chi^h = -\Psi\eta^h$, $\tau = 0, \tau^{b,h} = \tau^{X,h} - \chi\eta^h$, we get $\frac{\partial L}{\partial \chi}(\boldsymbol{\tau}, \chi) = \sum_h \left[\lambda\boldsymbol{\tau} - \lambda\boldsymbol{\tau}^{\xi,h} - \beta^h\boldsymbol{\tau}^{b,h}\right] \cdot \boldsymbol{c}_\chi^h = \lambda\sum_h \left[0 - \tau^{X,h} + \chi\eta^h\right]\Psi\eta^h$.

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 $\gamma^h = \lambda$.[35]

Proposition 4.4 *Assume $\gamma^h = \lambda$. Then jointly optimal nudges and taxes are given by the following formulas*

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

The more powerful the nudge is for high-internality agents (the higher is $\mathbb{E}\left[\tau^{X,h}\eta^h\right]$, 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 $\mathbb{E}\left[\tau^{X,h}m^h\right]$), 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 $\mathbb{E}\left[(m^h)^2\right]$, 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.[36] Similarly, the more heterogeneity there is in nudgeability (the higher is $\mathbb{E}\left[(\eta^h)^2\right]$, holding all other moments constant), then lower is the optimal nudge χ, and, under similar conditions, the higher is the optimal tax τ.

Nudges vs. taxes We now ask how to choose, if one must, between nudges and taxes. We could analyze this question using the model outlined just above, comparing the relative merits of nudges and taxes in terms of internality targeting and redistributive incidence. Instead, we choose to investigate this question in the context of a model with no heterogeneity, but where the nudges are potentially aversive.

We augment the example of Section 4.2 with aversive nudges. We use the same quadratic utility functions as in Section 4.3. We use the nudge as a tax model developed in Section 3.4. For concreteness, we interpret the harmful good (good 1) as cigarettes. We extend the model to account

[35] In the general case, with the notation $\sigma_{Y,Z} := cov\left(Y_h, Z_h\right)$:

$$\tau = \frac{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\lambda\tau^{X,h}m^h - \sigma_{\gamma,c/\Psi}\right] - \mathbb{E}\left[\gamma^h\eta_h m^h\right]\mathbb{E}\left[\lambda\tau^{X,h}\eta^h\right]}{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\gamma^h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\gamma^h\eta^h m^h\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]},$$
$$\chi = \frac{\mathbb{E}\left[\lambda\tau^{X,h}\eta^h\right]\mathbb{E}\left[\gamma_h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\lambda\tau^{X,h}m^h - \sigma_{\gamma,c/\Psi}\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]}{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\gamma^h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\gamma_h\eta^h m^h\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]}.$$

[36] The condition is $\mathbb{E}\left[\tau^{X,h}m^h\right]\mathbb{E}\left[{\eta^h}^2\right] \geq \mathbb{E}\left[\tau^{X,h}\eta^h\right]\mathbb{E}\left[\eta^h m^h\right]$. It is verified if η^h, m^h, $\tau^{X,h}$ are independent.

for the possibility that the nudge may directly create an aversive reaction (perhaps via a disgusting image of a cancerous lung), which we capture as a separable utility cost $\iota^h \chi c_i$ so that experienced utility is now

$$u^h(\boldsymbol{c}, \chi) = u^h(\boldsymbol{c}) - \iota^h \chi c_i,$$

where $\iota^h \chi c_i$ is the nudge aversion term. And we assume that there is no heterogeneity across agents.

The next proposition formalizes how nudge aversion changes the relative attractiveness of nudges vs. taxes. The planner must choose between two instruments to discourage cigarette consumption: a weakly positive tax ($\tau \geq 0$) or an aversive nudge ($\chi \geq 0$).

Proposition 4.5 ("Nudge the poor, tax the rich") *Consider a good with a "bad" internality (e.g. cigarettes). Suppose that at most one of two instruments (nudges and nonnegative taxes) can be used to correct this internality. An suppose that there is no heterogeneity across agents. Then an optimal tax is superior to an optimal nudge if and only if*

$$\frac{\lambda - \gamma^h}{m^h} > \frac{-\iota^h \gamma^h}{\eta^h}. \tag{28}$$

This proposition captures a new interesting trade-off between taxes and nudges. Both taxes and nudges correct internalities. But taxes also raise revenues on the agents consuming the good under consideration, which is desirable if $\lambda > \gamma^h$ but undesirable if $\lambda < \gamma^h$. Nudges do not raise revenues, and instead directly reduce utility.

When $\lambda > \gamma^h$, taxes dominate nudges as taxes have desirable side effects by raising revenues while nudges have adverse side effects by reducing utility. But when $\lambda < \gamma^h$ taxes and nudges both have undesirable side effects. Taxes dominate nudges when the desire to redistribute income towards agents consuming the good associated with the internality is weak ($\gamma^h - \lambda$ is low), and when these agents are attentive to the tax (m^h is high). Nudges dominate taxes when nudge aversion is low (ι^h is low) and when agents are easily nudged (η^h is high).

4.4 Do More Mistakes by the Poor Lead to More Redistribution?

In this example, we consider the consequences for income taxation of the hypothesis that richer agents make fewer mistakes (e.g., Mani, Mullainathan, Shafir and Zhao 2013). We deviate from the above setup with quasilinear utility. There are two goods with prices normalized to one. The decision and experienced utility of agent h are given respectively by

$$u^{s,h}(c_1, c_2) = \frac{c_1^{\alpha_1^{s,h}} c_2^{\alpha_2^{s,h}}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2}}, \qquad u^h(c_1, c_2) = \frac{c_1^{\alpha_1} c_2^{\alpha_2}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2}},$$

with $\alpha_1^{s,h} + \alpha_2^{s,h} = \alpha_1 + \alpha_2 = 1$. The indirect utility of agent h with (post-tax) income z is then given by

$$v^h(z) = A^h z, \qquad A^h = \left(\frac{\alpha_1^{s,h}}{\alpha_1}\right)^{\alpha_1} \left(\frac{\alpha_2^{s,h}}{\alpha_2}\right)^{\alpha_2} \leq 1.$$

The stronger the behavioral bias of agent h, the lower his marginal utility of income A^h. This is because a marginal unit of income is spent inefficiently.

We focus on a negative income tax: a linear income tax τ_z and a lump sum rebate. The unique purpose of the tax is to redistribute. The pre-tax and post-tax income of agent h are denoted by z^h and $z^h + \tau_z(\bar{z} - z^h)$, where $\bar{z}$ is average income defined by $\sum_h (z^h - \bar{z}) = 0$. We assume that the social welfare function is given by $\sum_h \frac{\left(v^h\right)^{1-\sigma}}{1-\sigma}$ with $\sigma > 0$.

The planning problem is then

$$\max_{\tau_Z} L(\tau_z; \{A^h\}),$$

with

$$L\left(\tau_z, \{A^h\}\right) = \sum_h \frac{\left(A^h\right)^{1-\sigma}}{1-\sigma} \left(z^h + \tau_z\left(\bar{z} - z^h\right)\right)^{1-\sigma}.$$

Proposition 4.6 (Mistakes and redistribution) *If the preference for redistribution is strong ($\sigma > 1$), larger behavioral biases (reductions in A^h) for the poor (agents with $z^h < \bar{z}$) lead to more redistribution (higher taxes τ_z). Conversely, if the preference for redistribution is weak ($\sigma < 1$), larger behavioral biases for the poor lead to less redistribution.*

The intuition for these results is the following. The key question is the impact of behavioral biases on the social marginal utility of income $\gamma^h = \left(A^h z\right)^{-\sigma} A^h$. Indeed, the marginal benefit $\frac{\partial L}{\partial \tau_z}$ of increased redistribution (a higher tax τ_z) is inversely related to the covariance between the social marginal utility of income and income $\frac{\partial L}{\partial \tau_z} = \sum_h \gamma^h \left(\bar{z} - z^h\right) = -cov\left(\gamma^h, z^h\right)$.

Larger behavioral biases for agent h increase its weight $\left(A^h z\right)^{-\sigma}$ in social welfare, but reduce his marginal utility of income A^h. The resulting effect on the social marginal utility of income γ^h depends on the relative strength of these two effects.[37] When $\sigma > 1$, the former effect dominates and larger behavioral biases lead to higher γ^h. The opposite occurs when $\sigma < 1$. When $\sigma = 1$, the two effects exactly cancel out so that γ^h is independent of A^h.

4.5 Mental Accounts

In this section, we show how our general model can accommodate some simple behavioral biases associated with mental accounting (Thaler 1985, Hastings and Shapiro 2013). We then show how these biases modify some basic results in optimal taxation.

[37] The NBER working paper version of Kaplow (2015) discusses a similar idea, in the context of a model with myopic agents.

An elementary model of mental accounts The primitives are an experienced utility function u, a partition of the set of commodities into K subsets or accounts indexed by $k = 1, ..., K$, mental accounting functions $\omega^k(\boldsymbol{q}, w)$, and an extended demand function $\boldsymbol{c}(\boldsymbol{q}, \boldsymbol{\omega})$, where $\boldsymbol{\omega} = \left(\omega^1, ..., \omega^K\right)$. We denote by $\boldsymbol{C}^k$ the vector of commodities associated with account k and we write $\boldsymbol{c} = \left(\boldsymbol{C}^1, .., \boldsymbol{C}^K\right)$. The mental accounting functions $\omega^k(\boldsymbol{q}, w)$ indicates how much money is devoted to account k, and must satisfy $\sum_k \omega^k(\boldsymbol{q}, w) = w$. The extended demand function must satisfy

$$\boldsymbol{q}^k \cdot \boldsymbol{C}^k(\boldsymbol{q}, \boldsymbol{\omega}) = \omega^k(\boldsymbol{q}, w).$$

The demand function $\boldsymbol{c}(\boldsymbol{q}, w)$ is simply defined by $\boldsymbol{c}(\boldsymbol{q}, w) = \boldsymbol{c}(\boldsymbol{q}, \boldsymbol{\omega}(\boldsymbol{q}, w))$. We denote the extended indirect utility function by $v(\boldsymbol{q}, \boldsymbol{\omega}) = u(\boldsymbol{c}(\boldsymbol{q}, \boldsymbol{\omega}))$.

Rational demand subject to mental accounts corresponds to

$$\boldsymbol{c}^r(\boldsymbol{q}, \boldsymbol{\omega}) = \arg\max_{\boldsymbol{c}} u(\boldsymbol{c}) \text{ s.t. } \boldsymbol{q}^k \cdot \boldsymbol{C}^k = \omega^k \text{ for } k = 1, ..., K. \tag{29}$$

The traditional model with frictionless mental accounts corresponds to $\omega^{k,r}(\boldsymbol{q}, w) = \boldsymbol{q}^k \cdot \boldsymbol{C}^{k,r}(\boldsymbol{q}, w)$, where $\boldsymbol{c}^r(\boldsymbol{q}, w) = \left(\boldsymbol{C}^{1,r}(\boldsymbol{q}, w), .., \boldsymbol{C}^{K,r}(\boldsymbol{q}, w)\right)$ is the demand function of a rational agent.

Given these mental accounting functions, spending on the different goods within each categories could be subject to additional behavioral biases. For example, a concrete model of the mental accounting functions $\omega^k(\boldsymbol{q}, w)$ is the following

$$\begin{aligned}(\boldsymbol{c}(\boldsymbol{q}, w), \boldsymbol{\omega}(\boldsymbol{q}, w)) = \arg \max_{\boldsymbol{c}=\left(\boldsymbol{C}^1, .., \boldsymbol{C}^K\right), \boldsymbol{\omega}} & u(\boldsymbol{c}) - \sum_k g^k\left(\omega^k - \omega_k^d(\boldsymbol{q}, w)\right) \\ \text{s.t. } \boldsymbol{q}^k \cdot \boldsymbol{C}^k = \omega^k \text{ for } k = 1, ..., K & \text{ and } \sum_k \omega^k = w,\end{aligned} \tag{30}$$

where $\omega_k^d(\boldsymbol{q}, w)$ is an exogenous default mental accounting function. The idea is that there are frictions on mental accounting so that the consumer faces some mental adjustment costs given by $g^k\left(\omega^k - \omega^{k,d}(\boldsymbol{q}, w)\right)$ when the expenditure $\omega^k(\boldsymbol{q}, w)$ on account k is different from the default expenditure $\omega^{k,d}(\boldsymbol{q}, w)$. Implicit in the formulation of this concrete model is the assumption that these mental adjustment costs are not taken into account in the evaluation of welfare. This approach is justifiable, but so is its polar alternative which fully takes these costs into account in welfare. Similar issues arise in the treatment of attention in Section 6.1. Note that the more abstract model presented above does not rely on this assumption.

The optimal tax formulas in Propositions 3.1 and 3.2 corresponding to the many-person Ramsey problem without and with externalities can then be applied without modifications to this simple model of mental accounting.

Rigid mental accounts We next explore concrete applications of this framework. In the interest of space, we focus in the main text on the case of rigid mental accounts, by which we

mean the following: a mental account k is *rigid* when the amount $\omega^k(\boldsymbol{q}, w)$ allocated to account k is independent of $\boldsymbol{q}$ and w (it must be the case that at least one account is not rigid, so that $\sum_{k'} \omega^{k'} = w$). We assume that the only bias is one of mental accounting, so that spending within each account is chosen to maximize utility. This means that the extended demand function is given by $\boldsymbol{c}(\boldsymbol{q}, \boldsymbol{\omega}) = \boldsymbol{c}^r(\boldsymbol{q}, \boldsymbol{\omega})$. The online appendix (Section 12.3) develops other concrete applications in the more general case with flexible accounts where $\omega^k(\boldsymbol{q}, w)$ depends on $\boldsymbol{q}$ and w.

Uniform commodity taxation with rigid mental accounts We first derive a uniform commodity taxation within rigid mental accounts.

Proposition 4.7 (Uniform commodity taxation within a rigid mental account) *Suppose that there is just one type of agent and that a mental account k is rigid. Then, all commodities associated with mental account k should be taxed at the same rate.*

It is efficient to tax all commodities associated in a rigid mental account at the same rate in order to avoid distorting the relative consumption of two commodities within the account.[38]

Rigid mental accounts therefore give a new, behavioral, rationale for uniform commodity taxes. It is distinct from the traditional argument for uniform commodity taxation with rational agents proposed by Atkinson and Stiglitz (1972), which obtains under some separability and homogeneity assumptions regarding preferences and which we review in Section 7.

Modified basic Ramsey and Pigou rules with rigid mental accounts We consider the basic setup in Section 2 with no misperceptions ($m_i = 1$ for all i) but with rigid mental accounts instead. We make the further simplification that there is one commodity per mental account. Consumption is therefore given by $c_i = \frac{\omega^i}{q_i} = \frac{\omega^i}{1+\tau_i}$. We assume that before taxes, the optimal amount ω^i is allocated to good i, so that $U^{i\prime}(\omega^i) = p_i = 1$, and that the rigid mental account ω^i does not adjust after the introduction of taxes.

We first derive the optimal Ramsey and Pigou tax rules with this rigid mental account with one good per account. Recall that we denote by $\psi_i = -\frac{U^{i\prime}(c_i)}{c_i U^{i\prime\prime}(c_i)}$ the inverse of the curvature of the utility function U^i for good i, which coincides with the demand elasticity of a rational agent.

Proposition 4.8 (Ramsey and Pigou formulas with rigid mental accounts) *Suppose that agents use a rigid mental account for good i, and the limit of small taxes. In the basic Ramsey problem, the optimal tax is*

$$\tau_i = \Lambda \psi_i \tag{31}$$

while in the basic Pigou problem, it is

$$\tau_i = \xi_i \psi_i. \tag{32}$$

[38] The Proposition implicitly assumes that all commodities in account k can be taxed.

The formula for the Ramsey problem is in stark contrast with the traditional Ramsey case where $\tau_i = \frac{\Lambda}{\psi_i}$, and the misperception case where $\tau_i = \frac{\Lambda}{m_i^2 \psi_i}$. With rigid mental accounts, a low (rational) elasticity ψ_i leads to low taxes, not to high taxes, as in the basic Ramsey. The intuition is as follows: if a good is very "necessary", rational demand is very inelastic: ψ_i is low. But with a rigid mental accounts, a tax τ_i leads to a consumption $c_i = \frac{\omega^i}{1+\tau_i}$. So, a high tax leads to a high distortion. Hence, when (rational) demand is very inelastic, the tax should be low.

Likewise, the modified Pigou formula $\tau_i = \xi_i \psi_i$ now features the rational elasticity of demand ψ_i. This is in contrast to the traditional case, where $\tau_i = \xi_i$, and to the case with misperception m_i where $\tau_i = \frac{\xi_i}{m_i}$ (Proposition 2.2).

To derive this result and understand it fully, it is useful to generalize it. We denote by α_i the elasticity of the demand for good i. In the traditional model without behavioral biases, we have $\alpha_i = \psi_i$. But in the model with attention m_i to the tax, we had $\alpha_i = m_i \psi_i$. With a rigid mental account for commodity i, given demand is $c_i = \frac{\omega^i}{1+\tau_i}$, the elasticity of the demand for good i is $\alpha_i = 1$. [39]

Proposition 4.9 (Ramsey and Pigou formulas with arbitrary behavioral elasticity) *Suppose that the rational demand elasticity for good is ψ_i, and that the behavioral demand elasticity is α_i. Consider the limit of small taxes. Then, in the basic Ramsey problem, the optimal tax is*

$$\tau_i = \Lambda \frac{\psi_i}{\alpha_i^2} \tag{33}$$

while in the basic Pigou problem, it is

$$\tau_i = \frac{\xi_i}{\alpha_i} \psi_i. \tag{34}$$

Proof. We could use the general formulas, but to gain intuition we proceed as follows, in the limit of small taxes. In the Ramsey problem, welfare can then be expressed as

$$L = -\frac{1}{2} \sum_i \frac{\alpha_i^2}{\psi_i} y_i \tau_i^2 + \Lambda \sum_i \tau_i y_i, \tag{35}$$

Indeed, a small tax τ_i changes consumption by $\delta c_i = -\alpha_i c_i \tau_i$. The associated distortion is $\frac{1}{2} (\delta c_i)^2 U^{i\prime\prime}(c_i) = \frac{1}{2} (-\alpha_i c_i \tau_i)^2 \frac{-U^{i'}(c_i)}{c_i \psi_i} = \frac{-1}{2} \frac{\alpha_i^2}{\psi_i} y_i \tau_i^2$ (recall that $\psi_i = -\frac{U^{i\prime}(c_i)}{c_i U^{i\prime\prime}(c_i)}$, and $U^{i\prime} = p_i = 1$ at the optimum, with $y_i = c_i$). Hence, the optimal tax is given by $L_{\tau_i} = 0$, i.e. $\tau_i = \Lambda \frac{\psi_i}{\alpha_i^2}$.

In the Pigou problem, at the first best, the planner would like $U^{i\prime}(c_i) = 1 + \xi_i$, as in the traditional tax. This means that consumption should change by $\delta c_i = -\psi_i \xi_i$ after the tax. But as the actual elasticity of demand is α_i, the tax should satisfy: $\delta c_i = -\alpha_i \tau_i = -\psi_i \xi_i$, and $\tau_i = \frac{\xi_i}{\alpha_i} \psi_i$. □

In the Ramsey problem, for a given demand elasticity α_i, a higher value of ψ_i pushes for higher

[39] Proposition 4.8 is a consequence of Proposition 4.9 when $\alpha_i = 1$ of the following result. Propositions 2.1 and 2.2 are also an application, when $\alpha_i = m_i \psi_i$.

tax, while for a given ψ_i, a higher value of α_i pushes for a lower tax. In the traditional model without behavioral biases, $\alpha_i = \psi_i$ and the resulting effect of a higher ψ_i is a lower tax. By contrast, in the behavioral model with a rigid mental account, $\alpha_i = 1$ so that a higher ψ_i results in a higher tax.

5 Nonlinear Income Taxation: Mirrlees Problem

5.1 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 $u_z \leq 0$.[40]

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", $\boldsymbol{Q} = (q(z))_{z \geq 0}$ 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 3. The primitives is the income function $z^n(q, \boldsymbol{Q}, r_0, r)$, which depends on the local marginal retention rate q, the ambient vector of all marginal retention rates $\boldsymbol{Q}$, and the virtual income r. In the traditional model without behavioral biases we have $z^n(q, \boldsymbol{Q}, r_0, r) = \arg\max_z u^n(qz + r, z)$, so that z^n does not depend on $\boldsymbol{Q}$ and r_0. With behavioral biases, this is no longer true in general. The income function is associated with the indirect utility function $v^n(q, \boldsymbol{Q}, r_0, r) = u^n(qz + r, z)_{|z = z^n(q, \boldsymbol{Q}, r_0, r)}$. The earnings $z(n)$ of agent n facing retention schedule $R(z)$ is then the solution of the fixed point problem $z = z^n(q(z), \boldsymbol{Q}, r(z))$. His consumption is $c(n) = R(z(n))$ and his utility is $v(n) = u^n(c(n), z(n))$.

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 $g(n) = W'(v(n)) v_r^n(q(z(n)), \boldsymbol{Q}, r(z(n)))$ the marginal utility of income. This is the analogue of β^h in the Ramsey problem of Section 3, 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).

[40] If the agent's pre-tax wage is n, L is his labor supply, and utility is $U^n(c, L)$, then $u^n(c, z) = U\left(c, \frac{z}{n}\right)$. Note that this assumes that the wage is constant (normalized to one). We discuss the impact of relaxing this assumption in Sections 7.1.2 and 12.7.

5.2 Saez Income Tax Formula with Behavioral Agents

5.2.1 Elasticity Concepts

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

$$\eta = q z_r(q, \boldsymbol{Q}, r_0, r).$$

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

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

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:

$$\zeta^c_{Q_{z^*}} = \frac{q}{z} z_{Q_{z^*}}(q, \boldsymbol{Q}, r_0, r).$$

We also define the earnings sensitivity to the lump-sum rebate at zero income[41]

$$\zeta^c_{r_0} = \frac{q}{z} z_{r_0}(q, \boldsymbol{Q}, r_0, r).$$

We shall call $\zeta^c_{Q_{z^*}}$ 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, $\zeta^c_{Q_{z^*}} = \zeta^c_{r_0} = 0$. But this is no longer true with behavioral agents.[42][43]

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 "misoptimization wedge"

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

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

[42] For instance, in the misperception model, in general, the marginal tax rate at z^* affects the default tax rate and therefore the perceived tax rate at earnings z.

[43] In the language of Section 3.1, we use income-compensation based notion of elasticity, $\boldsymbol{S}^C$, rather than the utility–compensation based notion $\boldsymbol{S}^H$.

We also define the renormalized misoptimization wedge

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

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

We have the following behavioral version of Roy's identity (proven in the online appendix, Section 13.3):

$$\frac{v_q}{v_w} = z - \frac{\tau^b z}{q}\zeta^c, \qquad \frac{v_{Q_{z^*}}}{v_w} = -\frac{\tau^b z}{q}\zeta^c_{Q_{z^*}}. \tag{36}$$

As in Section 3, the general model can be particularized to a decision vs. experienced utility model, or to a misperception model.

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

$$\max_{c,z|R^{s,n}(\cdot)} u^n(c,z) \text{ s.t. } c = R(z).$$

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

$$q^{s,n}(z)\,u^n_c(c,z) + u^n_z(c,z) = 0,$$

instead of the traditional condition $q(z)\,u^n_c(c,z) + u^n_z(c,z) = 0$. 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.[44]

Accordingly, we define $z^n(q, q^s, r)$ to be the solution of $q^{s,n}u^n_c(c,z) + u^n_z(c,z) = 0$ with $c = qz + r$.[45] The income $z(n)$ of agent n is then the solution of the fixed point equation

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

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

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

[44] This 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 2.1. The true constraint is $c = R(z)$, but the perceived constraint is $c = R^{s,n}(q, \boldsymbol{Q}, r_0)(z)$.

[45] If there are several solutions, we choose the one yield the greatest utility.

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

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

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. $q^d(\boldsymbol{Q}) = \int q(z)\omega(z)dz$ 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 $r' + q^s z^r(q^s, r') = r + qz^r(q^s, r')$. 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 $\zeta^{cr} = \frac{qS^r}{z}$ as the compensated elasticity of labor supply of the agent if he were rational.

We define $m_{zz} = q^s_q(q, \boldsymbol{Q}, r_0)(z)$ as the attention to the own marginal retention rate and $m_{zz^*} = q^s_{Q_{z^*}}(q, \boldsymbol{Q}, r_0)(z)$ 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 of the general model (the derivation is in Section 13.3 of the appendix):

$$\zeta^c = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{cr}m_{zz}, \qquad \zeta^c_{Q_{z^*}} = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{cr}m_{zz^*}, \tag{37}$$

$$\tau^b = \frac{\tau - \tau^s}{1 - \eta\frac{\tau-\tau^s}{q}}. \tag{38}$$

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 $\tau^b = 0$.

Decision vs. experienced utility model In the decision vs. experienced 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 $\zeta^c_{Q_{z^*}} = 0$, and ζ^c and η are the elasticities associated with decision utility u^s. The misoptimization wedge is

$$\tau^b = \frac{\frac{u_c}{u^s_c}u^s_z - u_z}{v_r}. \tag{39}$$

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

$$\gamma(z) = g(z) + \frac{\eta(z)}{1 - T'(z)} \left[\widetilde{\tau}^b(z) + \left(T'(z) - \widetilde{\tau}^b(z) \right) \frac{h^*(z)}{h(z)} \right]. \tag{40}$$

This definition is the analogue of the corresponding definition (14) in the Ramsey model. It is motivated by Lemma 13.2 in the online appendix, which shows that, 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 $\frac{\eta(z)}{1-T'(z)}\widetilde{\tau}^b(z)\delta K$; 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 misoptimization.

5.2.2 Optimal Income Tax Formula

We next present the optimal income tax formula. The online appendix (section 13.2) presents the intermediary steps used in the derivation of this formula.

Proposition 5.1 *Optimal taxes satisfy the following formulas (for all z^*)*

$$\begin{aligned} \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} = & \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} (1 - \gamma(z)) \frac{h(z)}{1 - H(z^*)} dz \\ & - \int_0^{\infty} \frac{\zeta^c_{Q_{z^*}}(z)}{\zeta^c(z^*)} \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} \frac{z h^*(z)}{z^* h^*(z^*)} dz. \end{aligned} \tag{41}$$

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

$$\begin{aligned} & \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} + \int_0^{\infty} \omega(z^*, z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} dz \\ & = \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} e^{-\int_{z^*}^{z} \rho(s) ds} \left(1 - g(z) - \eta \frac{\tilde{\tau}^b(z)}{1 - T'(z)} \right) \frac{h(z)}{1 - H(z^*)} dz, \end{aligned} \tag{42}$$

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

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

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 (41) 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 (40)). The second term $-\frac{1}{z^*}\int_0^\infty \frac{\zeta^c_{Q_{z^*}}(z)}{\zeta^c(z^*)}\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}z\frac{h^*(z)}{h^*(z^*)}dz$ 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 $\zeta^c_{Q_{z^*}}(z) > 0$, 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 (42) 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 (41) using equation (40). When there are no income effects so that $\eta = \rho(z) = 0$, the optimal tax formula (41) and the modified Saez formula (42) are identical. They coincide with the traditional Saez formula when there are no behavioral biases so that $\zeta^c_{Q_{z^*}}(z) = \omega(z^*, z) = \tilde{\tau}^b(z) = 0$. In this case, the left-hand side of (42) 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 additional sufficient statistic, namely the misoptimization wedge $\tilde{\tau}^b(z)$ and the behavioral cross-elasticities $\zeta^c_{Q_{z^*}}(z)$. Second, it is not possible to solve out the optimal marginal tax rate in closed form. Instead, the modified Saez formula (42) 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 decision vs. experienced utility model. Indeed, we then have $\omega(z^*, z) = 0$ and $\tilde{\tau}^b(z) = g(z)\frac{u_c\frac{u^s_z}{u^s_c} - u_z}{v_r}$, so that there is no linear system of equations to solve out to recover $T'(z)$.

5.2.3 Marginal Tax Rate for Top Incomes

We start by revisiting the classic result that if the income distribution is bounded at $z_{\max}$, then the top marginal income tax rate should be zero. In our model, this need not be the case. One simple way to see that is to consider the case of decision vs. experienced utility. Tax formula (41) then prescribes $T'(z_{\max}) = \tilde{\tau}^b(z_{\max})$ which is positive or negative depending on whether top earners over or under perceive the benefits of work (under or over perceive the costs of work).

We now derive a formula for the marginal rate at very high incomes when the income distribution is unbounded at the top. It proves convenient to consider a (high) z_0 above which we consider that incomes are "top incomes", and the marginal rate is constant. We consider tax systems with constant marginal tax rates for $z \geq z_0$. We assume that $g(z) = \overline{g}$ for $z > z_0$. We call $\zeta^c_{\overline{q}}(z) = \int_{z_0}^\infty \zeta^c_{Q_{z^*}}(z)dz_*$ the sensitivity to the asymptotic tax rate. This is the elasticity of earnings

of an individual at earnings $z < z_0$ to an increase to the top rate, arising perhaps because of a misperception of the tax environment. Concretely, think of the recent case of France where increasing the top rate to 75% might have created an adverse general climate with the perception that even earners the top income would pay higher taxes.

We call $\overline{\eta}$, $\overline{g}, \overline{\zeta}^c$ the asymptotic values for large incomes and π the Pareto exponent of the earnings distribution (when z is large, $1-H(z) \propto z^{-\pi}$). We define the weighted means: $\mathbb{E}^z[\phi(z)] = \frac{\int \phi(z)h(z)zdz}{\int \phi(z)zdz}$ and $\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right] = \frac{\int \zeta^c_{\overline{q}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}h^*(z)dz}{\int \frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}h^*(z)dz}$.

Proposition 5.2 (Optimal tax rate for top incomes) *The optimal marginal rate $\overline{\tau}$ for top incomes is*

$$\overline{\tau} = \frac{1-\overline{g}-\beta+\overline{\zeta}^c\pi\overline{g}\tau^b}{1-\overline{g}-\beta+\overline{\zeta}^c\pi+\overline{\eta}} \tag{43}$$

where

$$\beta = \mathbb{E}^z\left[\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}\right]\pi\frac{\mathbb{E}[z]}{\mathbb{E}[z1_{z\geq z_0}]}\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right].$$

This generalizes the Saez (2001) formula which can be recovered in the particular case where $\beta = \tau^b = 0$. The intuition is as follows—the β terms reflects not only the fact that the top marginal tax rate affects not only top earners, but also the tax perceived by agents at all points of the income distribution with associated effects on tax revenues. The more increasing the top tax rate lowers all incomes (the higher $\zeta^c_{\overline{q}}(z)$), the higher β, and the lower the top optimal tax rate.

The τ^b terms are positive (resp. negative) when top earners overperceive (resp. underperceive) the marginal benefits of effort or underperceive (resp. overperceive) taxes. These terms lead to higher (resp. lower) top optimal rates compared to the Saez formula.

Consider the typical Saez calibration with $\zeta^c(\infty) = 0.2$, $\eta = 0$ and $\pi = 2$. If the typical tax is $T'(z) \simeq \frac{1}{3}$ so that $\mathbb{E}^z\left[\frac{T'(z)}{1-T'(z)}\right] \simeq \frac{1}{2}$, we take z_0 to be at the top 1% quantile of the income distribution. Piketty and Saez (2003, updated 2015) report that in the income share of the top 1% is 20%, so that $\frac{\mathbb{E}[z]}{\mathbb{E}\left[z1_{z\geq z_0}\right]} = \frac{1}{0.2}$. This implies that $\beta = \mathbb{E}^z\left[\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}\right]\pi\frac{\mathbb{E}[z]}{\mathbb{E}\left[z1_{z\geq z_0}\right]}\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right] = \frac{1}{2}2\frac{1}{0.2}\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right] = 5E^*\left[\zeta^c_{\overline{q}}\right]$. Also, we take top earnings to be well calibrated, i.e. $\tau^b = 0$.

The average cross-influence $\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right]$ does not appear to have ever been measured. It is assumed to be 0 in the traditional model. We propose the following thought experiment to gauge its potential magnitude. Suppose that increasing the top rate by 10% will decrease earnings outside the top bracket by $x = 1\%$. Then, $\mathbb{E}^*\left[\zeta^c_{\overline{q}}\right] = (1-T'(z))\frac{z_{\overline{q}}}{z} = \left(1-\frac{1}{3}\right)\frac{x}{0.1} = 6.7x$, which gives an interpretable benchmark that we now use.

Take first the case where $\overline{g} = 0$ the top optimal tax rate maximizes revenues raised on top earners. With rational agents ($x = 0$), the top marginal tax rate is $\overline{\tau} = 71\%$. If $x = 1\%$, then $\overline{\tau} = 62\%$, and if $x = 2\%$, then $\overline{\tau} = 45\%$. If $x = -1\%$, then $\overline{\tau} = 77\%$.[46] When the weight on top

[46] We thank Thomas Piketty for suggesting to us that if workers are happier, and strike less, because the taxes on the wealth has high, then $x < 0$.

earnings is higher, say $\overline{g} = 0.2$, the corresponding numbers for the top rate are: 67%, 53% and 25%, and 74%. This illustrates the potentially large importance of the behavioral cross-impact of the top tax rate, a sufficient statistic that is assumed to be zero in traditional analyses.

The misoptimization wedge τ^b does not affect the optimal tax rate when $\overline{g} = 0$. When $\overline{g} = 0.5$, it increases the optimal top rate from 56% to 67% when the internality goes from no misperception of taxes by top earners $\tau^b = 0$ to underperception of taxes by top earners $\tau^b = 0.5$.

5.2.4 Possibility of Negative Marginal Income Tax Rates

In the traditional model with no behavioral biases, negative marginal income tax rates can never arise at the optimum. With behavioral biases negative marginal income tax rates are possible at the optimum. To see this, consider for example the decision vs. experienced utility model with decision utility u^s and assume that u^s is quasilinear so that there are no income effects $u^s(c,z) = c - \phi\varkappa(z)$. We take experienced utility to be $u(c,z) = \theta c - \phi(z)$. Then the modified Saez formula (42) becomes

$$\frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} = \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} (1 - g(z)) \frac{h(z)}{1 - H(z^*)} dz,$$

where $\tilde{\tau}^b(z) = -g(z)\phi'(z)\frac{\theta - 1}{\theta}$ by (39). When $\theta > 1$, we have $\widetilde{\tau}^b(z^*) < 0$, and it is possible for this formula to yield $T'(z^*) < 0$. This occurs if agents undervalue the benefits or overvalue the costs from higher labor supply. For example, it could be the case that working more leads to higher human capital accumulation and higher future wages, but that these benefits are underperceived by agents, which could be captured in reduced form by $\theta > 1$. Such biases could be particularly relevant at the bottom of the income distribution (see Chetty and Saez (2013) for a review of the evidence). If these biases are strong enough, the modified Saez formula could predict negative marginal income tax rates at the bottom of the income distribution. This could provide a behavioral rationale for the EITC program.[47] In parallel and independent work, Gerritsen (2014) and Lockwood (2015) derive a modified Saez formula in the context of decision vs. experienced utility model. Lockwood (2015) zooms in on the EITC program and provides an empirical analysis documenting significant present-bias among EITC recipients and shows that a calibrated version of the model goes a long way towards rationalizing the negative marginal tax rates associated with the EITC program.

This differs from alternative rationales for negative marginal income tax rates that have been put forth in the traditional literature. For example, Saez (2002) shows that if the Mirrlees model is extended to allow for an extensive margin of labor supply, then negative marginal income tax rates can arise at the optimum. We refer the reader to the online appendix (section 12.6) for a behavioral treatment of the Saez (2002) extensive margin of labor supply model.

[47] The EITC program itself could be misperceived, see Chetty, Friedman and Saez (2013).

6 Endogenous Attention and Salience

We now allow for endogenous attention to taxes and analyze its impact on optimal taxes. We also discuss tax salience as a policy choice in the design of the optimal tax system. We illustrate the discussion in the context of the general analysis of Section 3.

6.1 Attention and Welfare

Attention as a good To capture attention and its costs, we propose the following reinterpretation of the general framework. We imagine that we have the decomposition $\boldsymbol{c} = (\boldsymbol{C}, \boldsymbol{m})$, where $\boldsymbol{C}$ is the vector of traditional goods (champagne, leisure), and $\boldsymbol{m}$ is the vector of attention (e.g. m_i is attention to good i). We call $I^{\boldsymbol{C}}$ (respectively I^m) the set of indices corresponding to traditional goods (respectively attention). Then, all the analysis and propositions apply without modification. We here summarize the essentials, while Section 12.2 in the appendix gives more details.

This flexible modeling strategy allows to capture many potential interesting features of attention. The framework allows (but does not require) attention to be chosen and to react endogenously to incentives in a general way (optimally or not). It also allows (but does not require) attention to be produced, purchased and taxed.

We find it most natural to consider the case where attention is not produced, cannot be purchased, and cannot be taxed. This case can be captured in the model by imposing that ($p_i = \tau_i = 0$ for $i \in I^m$).

It is useful to consider two benchmarks. The first benchmark is "optimally allocated attention", which we capture as follows: we suppose that there is a primitive choice function $\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m})$ for traditional goods that depends on attention $\boldsymbol{m}= (m_1, ..., m_A)$ so that $\boldsymbol{c}(\boldsymbol{q}, w, \boldsymbol{m}) = (\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m}), \boldsymbol{m})$.[48] Attention $\boldsymbol{m}= \boldsymbol{m}(\boldsymbol{q}, w)$ is then chosen to maximize $u(\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m}), \boldsymbol{m})$. This generates a function $\boldsymbol{c}(\boldsymbol{q}, w) = (\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m}(\boldsymbol{q}, w)), \boldsymbol{m}(\boldsymbol{q}, w))$. In that benchmark, attention costs are incorporated in welfare.[49] For instance we might consider a separable utility function $u(\boldsymbol{C}, \boldsymbol{m}) = U(\boldsymbol{C}) - g(\boldsymbol{m})$ for some cost function $g(\boldsymbol{m})$. A non-separable u might capture that attention is affected by consumption (e.g., of coffee) and attention affects consumption (by needing aspirin).

The second benchmark is "no attention cost in welfare," where attention is endogenous (given by a function $\boldsymbol{m}(\boldsymbol{q}, w)$) but its cost is assumed not to directly affect welfare so that $u(\boldsymbol{C}, \boldsymbol{m}) = U(\boldsymbol{C})$. For instance, as a decision vs. experienced utility generalization of the example of the previous paragraph, we could have $\boldsymbol{m}(\boldsymbol{q}, w) = \arg\max_{\boldsymbol{m}} u^s(\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m}), \boldsymbol{m})$, where $u^s(\boldsymbol{C}, \boldsymbol{m}) = U(\boldsymbol{C}) - g(\boldsymbol{m})$, but still $u(\boldsymbol{C}, \boldsymbol{m}) = U(\boldsymbol{C})$. In that view, people use decisions heuristics that can respond

[48] For instance, in a misperception model, attention operates by changing the perceived price $\boldsymbol{q}^s(\boldsymbol{q}, w, \boldsymbol{m})$ which in turn changes consumption as $\boldsymbol{C}(\boldsymbol{q}, w, \boldsymbol{m}) = \boldsymbol{C}^s(\boldsymbol{q}, \boldsymbol{q}^s(\boldsymbol{q}, w, \boldsymbol{m}), w)$.

[49] The first order condition characterizing the optimal allocation of attention can be written as $\boldsymbol{\tau}^b \cdot \boldsymbol{c}_{m_j}(\boldsymbol{q}, w, \boldsymbol{m}) = 0$ for all $j \in \{1, ..., A\}$. This condition can be re-expressed more conveniently by introducing the following notation: we call $k(i)$ the index $k \in I^m$ corresponding to dimension $i \in \{1, ..., A\}$ of attention. We then get $\sum_{i \in I^C} \tau_i^b \boldsymbol{C}_{m_j}(\boldsymbol{q}, w, \boldsymbol{m}) + \tau_{k(j)}^b = 0$ for all $j \in \{1, ..., A\}$.

to incentives, but the cost of those decision heuristics is not counted in the utility function. In this benchmark, we have $\tau_i^b = 0$ for $i \in I^m$.

Optimal taxation with endogenous attention The tax formula (17) has a term $(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{S}_i^{C,h} = \sum_{k \in I^m \bigcup I^c} (\tau_k - \widetilde{\tau}_k^{b,h}) \boldsymbol{S}_{ki}^{C,h}$, a sum that includes the "attention" goods $k \in I^m$. As attention is assumed to have zero tax, we have $\tau_k = 0$ for $k \in I^m$. The term $\widetilde{\tau}_k^{b,h}$, which accounts for potential misoptimization in the allocation of attention, requires no special treatment. However, two polar special cases are worth considering that simplify the calculations.

First, consider the "no attention cost in welfare" case. In this case we saw that $\widetilde{\tau}_k^{b,h} = 0$ for $k \in I^m$. Together with $\tau_k = 0$ for $k \in I^m$, this implies that $(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{S}_i^{C,h} = \sum_{k \in I^C} (\tau_k - \widetilde{\tau}_k^{b,h}) S_{ki}^{C,h}$ is the sum restricted to commodities.

Second, consider the "optimally allocated attention" case. Then (see Proposition 12.3 in the online appendix)

$$(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{S}_i^{C,h} = \sum_{k \in I^C} (\tau_k S_{ki}^{C,h} - \widetilde{\tau}_k^{b,h} S_{ki|m}^{C,h}) \tag{44}$$

where $\boldsymbol{S}_{i|\boldsymbol{m}}^{C,h}$ is a Slutksy matrix holding attention constant, which is in general different from $\boldsymbol{S}_i^{C,h}$. For tax revenues, the full Slutsky matrix, including changes in attention, matters (the term $\tau_k S_{ki}^{C,h}$). However, for welfare, when attention is assumed to be optimally allocated, it is the Slutsky matrix holding attention constant that matters (the term $\widetilde{\tau}_k^{b,h} S_{ki|\boldsymbol{m}}^{C,h}$). This is a version of the envelope theorem.

Illustration in the basic Ramsey case We illustrate these notions in the basic Ramsey case of section 2 with just one taxed good (good 1, whose index we drop, and whose pre-tax price is 1), in the limit of small taxes. We go back to a very elementary presentation, as this provides a clear intuition for the economic forces at work. 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

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

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

$$L'(\tau) = -\psi y m(\tau)\tau\left(m(\tau) + \tau m'(\tau)\right) - Ag'\left(m(\tau)\right)m'(\tau) + \Lambda y = 0.$$

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

$$\tau^{m,*} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m'(\tau)} \tag{45}$$

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

$$\tau^{m,0} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m(\tau) m'(\tau)} \tag{46}$$

When attention is fixed, the optimal tax is

$$\tau^{m,F} = \frac{\Lambda/\psi}{m(\tau)^2}. \tag{47}$$

Proposition 6.1 *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. [50] Plainly, the tax is higher when attention is variable than when attention is fixed—this is basically because demand is more elastic then $(-\frac{p}{c}\frac{\partial c}{\partial \tau} = -\psi\left(m(\tau) + \tau m'(\tau)\right))$.

6.2 Salience as a Policy Choice

Governments have a variety of ways of making a particular tax more or less salient. For example, Chetty, Kroft and Looney (2009) present evidence that sales taxes that are included in the posted prices that consumers see when shopping have larger effects on demand. It is therefore not unreasonable to think of salience as a characteristic of the tax system that can be chosen or at least influenced by the government. This begs the natural question of the optimal salience of the tax system.[51]

[50] The example allows to appreciate the Slutsky matrix with or without constant attention. The Slutsky matrix with constant m has $S^C_{11|m} = \frac{\partial c(1+\tau,m)}{\partial \tau} = -\psi c m$, while the Slutksy matrix with variable m has $S^C_{11} = \frac{dc(1+\tau,m(\tau))}{d\tau} = -\psi c\left(m + \tau m'(\tau_1)\right)$. The online appendix (section 12.2.2) provides other illustrations.

[51] Note that we are excluding taxes from directly affecting experienced utility: taxes affect utility only through their impact on the chosen consumption bundle. In the terminology of Bernheim and Rangel (2009), we treat tax

We investigate this question in the context of two simple examples. We start with the basic Ramsey model developed in Section 2. Imagine that the government can choose between two tax systems with different degrees of salience m and m' with $m'_i > m_i$ for all i. Then it is optimal for the government to choose the lowest degree of salience. We denote by L (respectively L') the value of the objective of the government with optimal taxes conditional on salience $\boldsymbol{m}$ (respectively $\boldsymbol{m}'$). The gain from decreased salience can be written as $L - L' = \Delta L$ with $\Delta L = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{m_i^2} - \frac{1}{m_i'^2}\right)\frac{\Lambda^2}{\psi_i}y_i > 0$. In this basic Ramsey model where taxes are used only to raise tax revenues, less salient taxes are preferable. The reason is that less salient taxes are less distortionary.

The result that less salience is desirable is not general. We choose to make this point in the context of the basic Pigou model developed in Proposition 2.3. We suppose that all agents have the same utility $U^h = U$ and the same associated externality/internality $\xi^h = \xi$, but that their perceptions are different $m^h \neq m^{h'}$. Now imagine that the government can choose between two tax systems with different degrees of salience m^h and m'^h for every agent h. We want to capture the idea that more salience not only increases average attention for all agents, but also decreases the heterogeneity in perceptions . We formalize this comparative static with the following two requirements $\mathbb{E}[m'^h] < \mathbb{E}[m^h]$ and $\frac{Var[m^h]}{\mathbb{E}[m^{h2}]} < \frac{Var[m^{h'}]}{\mathbb{E}[m^{h'2}]}$. The gain from the decreased salience is given by $L - L' = \Delta L$ with $\Delta L = -\frac{1}{2}\Psi H \xi^2\left(\frac{Var[m^h]}{\mathbb{E}[m^{h2}]} - \frac{Var[m^{h'}]}{\mathbb{E}[m^{h'2}]}\right) < 0$. In this basic Pigou model, where taxes are used only to correct a homogenous externality/internality, more salient taxes are preferable. This is because more salient taxes are perceived more homogeneously and can therefore better correct for the underlying externality/internality.

In general the key observation is that the relevant Slutsky matrix $\boldsymbol{S}_i^{C,h}$ that appears in the optimal tax formula depends on the salience of the tax system. It could also be interesting to allow the government to combine different tax instruments with the same tax base but different degrees of salience. Our general model could be extended to allow for this possibility. We would start with a function $c\left(w, \boldsymbol{p}, \boldsymbol{\tau}^1, \boldsymbol{\tau}^2, ..., \boldsymbol{\tau}^K\right)$, where $\boldsymbol{\tau}^\kappa$ are tax vectors with different degrees of salience. Each tax instrument κ corresponds to a Slutsky matrix $S_{ij}^{C,\kappa}$ which depends on the tax instrument indexed by κ. The optimal tax formula can then be written as $\frac{\partial L(\boldsymbol{\tau})}{\partial \tau_i^\kappa} = 0$ where

$$\frac{\partial L\left(\boldsymbol{\tau}\right)}{\partial \tau_i^\kappa} = \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \lambda(\bar{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot \boldsymbol{S}_i^{C,\kappa,h}],$$

with $\bar{\boldsymbol{\tau}} = \sum_{\kappa=1}^{K} \boldsymbol{\tau}^\kappa$. The intuition for this formula is that the different tax instruments lead to different substitution effects captured by different Slutsly matrices $S_{ij}^{C,\kappa}$. Note that the substitution effect associated with one tax instrument κ affects the common tax base of the other tax instruments κ', which explains why the formula features the total $\bar{\boldsymbol{\tau}}$ rather than the individual tax $\boldsymbol{\tau}^\kappa$.[52]

salience as an anciliary condition.

[52] 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

7 Revisiting Diamond-Mirrlees and Atkinson-Stiglitz

In this section, we revisit two classical public finance results: the Diamond and Mirrlees (1971) production efficiency result and the associated result that supply elasticities do not enter in optimal tax formulas, as well as the Atkinson and Stiglitz (1972) uniform commodity taxation result.

7.1 Diamond-Mirrlees (1971)

7.1.1 Supply Elasticities: Optimal taxes with Efficient Production

So far, we have assumed a perfectly elastic production function (constant production prices $\boldsymbol{p}$). In traditional, non-behavioral models, this is without loss of generality. Indeed, with a complete set of commodity taxes, optimal tax formulas depend only on production prices but not on production elasticities $\boldsymbol{\tau}$. In behavioral models, this result must be qualified. This section therefore generalizes the model to imperfectly elastic production function (non-constant prices $\boldsymbol{p}$).

In behavioral models, prices $\boldsymbol{p}$ and taxes $\boldsymbol{\tau}$ might affect behavior differently. We introduce a distinction between taxes $\boldsymbol{\tau^p}$, that affect behavior like prices, and taxes, $\boldsymbol{\tau}^c$ that affect behavior different from prices. For example, $\boldsymbol{\tau^p}$ could represent taxes that included in listed prices $\boldsymbol{p}+\boldsymbol{\tau^p}$ (either because they are levied on producers or because they are levied on consumers but the listed prices are inclusive of the tax) and taxes $\boldsymbol{\tau}^c$ that are not included in listed prices. An agent's demand function can then be written as $\boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau^p},\boldsymbol{\tau}^c,w\right)$. This distinction will prove to be important for the generalization of our results to imperfectly elastic production functions.

We denote the associated indirect utility function by $v^h\left(\boldsymbol{p}+\boldsymbol{\tau^p},\boldsymbol{\tau}^c,w\right)$ and the Slutsky matrices corresponding to $\boldsymbol{\tau}^p$ (or $\boldsymbol{p}$) and $\boldsymbol{\tau}^c$by $\boldsymbol{S}_i^{H,\boldsymbol{p},h}$ and $\boldsymbol{S}_i^{H,c,h}$, respectively. We allow for the possibility (but we do not impose it) that these Slutsky matrices do not coincide.

We assume that government must finance a vector of government consumption $\boldsymbol{g}$ and that profits are fully taxed—we allow for decreasing returns to scale and nonzero profits. The production set is expressed as $\{\boldsymbol{y}$ s.t. $F\left(\boldsymbol{y}\right)\leq 0\}$. Perfect competition imposes that $F\left(\boldsymbol{y}\right)=0$ and $\boldsymbol{p}=F'\left(\boldsymbol{y}\right)$, where $\boldsymbol{y}$ is the equilibrium production. Market clearing requires that $\boldsymbol{g}+\sum_h \boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)=\boldsymbol{y}$. We denote by $\bar{\boldsymbol{\tau}}=\boldsymbol{\tau}^c+\boldsymbol{\tau}^p$ the sum of the tax vectors.

(taxed) good. With $m'>m$, we get

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

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 (2012). In general, with more than one agent and heterogeneities in the mispereceptions of the two taxes, the first best might not be achievable.

We can write the planning problem as

$$\max_{\boldsymbol{p},\boldsymbol{\tau}^p,\boldsymbol{\tau}^c} W\left(\left(v^h\left(\boldsymbol{p}+\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)\right)_{h=1...H}\right)$$

subject to the resource constraint

$$F\left(\boldsymbol{g}+\sum_h \boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)\right)=0,$$

and the competitive pricing condition

$$\boldsymbol{p}=F'\left(\boldsymbol{g}+\sum_h \boldsymbol{c}^h\left(\boldsymbol{p}+\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)\right).$$

The competitive pricing equation is a fixed point in $\boldsymbol{p}$. We denote the solution by $\boldsymbol{p}\left(\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)$. The derivatives of this function $\boldsymbol{p}$ encapsulate the incidence of taxes depending on the demand and supply elasticities. We define the price derivative matrix by $\varepsilon_{ij}^{\kappa}=\frac{\partial p_i}{\partial \tau_j^{\kappa}}$, the derivative of the prices p_i of commodity i with respect to the tax τ_j^{κ} with $\kappa\in\{p,c\}$. We also define the supply elasticity matrix ε_S by $\varepsilon_{S,ij}=\frac{p_j}{y_i}\left(F''^{-1}\right)_{ij}$ and the demand elasticities ε_D^{κ} by $\varepsilon_{D,ij}^{\kappa}=-\frac{1}{y_i}\sum_h p_j \boldsymbol{c}_{i,\tau_j^{\kappa}}^h$. Finally we define the matrix $diag\left(\boldsymbol{p}\right)$ as the diagonal matrix with i-th element given by p_i. Then, by applying the implicit function theorem to the competitive pricing condition, we obtain after some manipulations that the matrix of price derivatives ε^{κ} is given by

$$\varepsilon^{\kappa}=-diag(\boldsymbol{p})\left(\varepsilon_S+\varepsilon_D^p\right)^{-1}\varepsilon_D^{\kappa}diag\left(\boldsymbol{p}\right)^{-1} \tag{48}$$

so that the ε^{κ} reflects both demand and supply elasticities. This formula is the behavioral extension of the standard incidence calculations determining how the burden of taxes is shared between consumers and producers. Compared with the traditional model without behavioral biases, the difference is that ε_D^{κ} depends on the salience of the tax instrument κ. Incidence ε^{κ} therefore depends on salience (and more generally on how taxes are perceived). This conceptual point already appears in Chetty, Looney and Kroft (2009). Our incidence formula only generalizes it to many goods and arbitrary preferences.

We replace $\boldsymbol{p}$ in the objective function and the resources constraint, and we put a multiplier λ on the resource constraint. We form the Lagrangian

$$L\left(\boldsymbol{\tau}^p,\boldsymbol{\tau}^c\right)=W\left(\left(v^h\left(\boldsymbol{p}\left(\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)+\boldsymbol{\tau}^{\boldsymbol{p}},\boldsymbol{\tau}^c,w\right)\right)_{h=1...H}\right)-\lambda F\left(\boldsymbol{g}+\sum_h \boldsymbol{c}^h\left(\boldsymbol{p}\left(\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)+\boldsymbol{\tau}^p,\boldsymbol{\tau}^c,w\right)\right).$$

The optimal tax formulas can be written as $L_{\tau_i^{\kappa}}=0$ for $\kappa\in\{p,c\}$ if tax τ_i^{κ} is available.

Proposition 7.1 *With an imperfectly elastic production function, the following results hold. First,*

if there is a full set of commodity taxes $\boldsymbol{\tau}^p$, then the optimal tax formulas can be written as

$$0=\sum_h[\left(\lambda-\gamma^h\right)c_i^h+\lambda(\bar{\boldsymbol{\tau}}-\widetilde{\boldsymbol{\tau}}^{b,h})\cdot\boldsymbol{S}_i^{H,\kappa,h}]$$

and are independent of production elasticities and coincide with those of Section 3 if taxes are restricted to be of the $\boldsymbol{\tau}^p$ type or with those of Section 6.2 if taxes can be both of the $\boldsymbol{\tau}^p$ type and the $\boldsymbol{\tau}^c$ type. Second, when there is a restricted set of commodity taxes $\boldsymbol{\tau}^p$, then the optimal tax formulas can be written as

$$0=\sum_h[\left(\lambda-\gamma^h\right)c_i^h+\lambda(\bar{\boldsymbol{\tau}}-\widetilde{\boldsymbol{\tau}}^{b,h})\cdot\boldsymbol{S}_i^{H,\kappa,h}]+\sum_h\sum_j[\left(\lambda-\gamma^h\right)c_j^h+\lambda(\bar{\boldsymbol{\tau}}-\widetilde{\boldsymbol{\tau}}^{b,h})\cdot\boldsymbol{S}_j^{H,p,h}]\varepsilon_{ji}^{\kappa}$$

which depend on production elasticities and do not coincide with those of Sections 3 or 6.2.

With a full set of commodity taxes $\boldsymbol{\tau}^p$, we can rewrite the objective function and the resource constraint in the planning problem as a function of $\boldsymbol{q}=\boldsymbol{p}+\boldsymbol{\tau}^p$. We can then relax the planning problem by dropping the competitive pricing equation, which is slack—this equation can then simply be used to find $\boldsymbol{\tau}^p$ given the desired value of $\boldsymbol{q}$. As a result, only the first derivatives of the production function $\boldsymbol{p}=F'$ enter the optimal tax formulas and not the second derivatives F'' (and hence do not depend on supply elasticities). With a restricted set of commodity taxes $\boldsymbol{\tau^p}$, this relaxation of the planning problem fails, the competitive pricing equation cannot be dropped, and the optimal tax formulas depend on the second derivatives F'' (and hence depend on supply elasticities).

Therefore, with behavioral agents, the principle from traditional models that supply elasticities do not enter in optimal tax formulas as long as there is a full set of commodity taxes extends if taxes are understood to be of the $\boldsymbol{\tau}^p$ form. The difference is that even with a full set of commodity taxes of the $\boldsymbol{\tau}^c$ type (which would be enough to guarantee that optimal tax formulas do not depend on supply elasticities in the traditional model), optimal tax formulas do depend on supply elasticities if there is only a restricted set of commodity taxes of the $\boldsymbol{\tau}^p$ form.

A similar result holds in the Mirrlees case. Hence, in the traditional analysis, the supply elasticity doesn't appear in the optimal tax formula. This is not true any more with a behavioral model, which is developed in Proposition 12.14 of the online appendix.

To illustrate Proposition 7.1, consider the separable case $u\left(\boldsymbol{c}\right)=c_0+u\left(c_1\right)$ in the misperception case with $\tau_1^s=\tau_1^p+m_1\tau_1^c$, $0\leq m_1\leq 1$ and τ_1^p is exogenous (perhaps set to 0).

We represent the production function as follows—it takes $C\left(y_1\right)$ units of good 0 to produce y_1 units of good 1. We define supply and demand to be $S\left(p_1\right)=C'^{-1}\left(p_1\right)$ and $D\left(p_1+\tau_1^p+m_1\tau_1^c\right)=u'^{-1}\left(p_1+\tau_1^p+m_1\tau_1^c\right)$. We denote the corresponding supply and demand elasticities (corresponding to a fully perceived change in p_1) by $\varepsilon_S>0$ and $\varepsilon_D>0$. Differentiating the equilibrium condition

$S(p_1) = D(p_1 + \tau_1^p + m_1\tau_1^c)$ yields

$$\varepsilon_{11}^c = -\frac{\varepsilon_D}{\varepsilon_S + \varepsilon_D} m_1.$$

Compared to the traditional incidence analysis, because consumers are not fully attentive to the tax on good 1 ($m_1 < 1$), the burden of the tax is shifted to the consumer. This echoes a result in Chetty, Looney and Kroft (2009).

We now turn to optimal taxes. We work in the limit of small taxes when $\Lambda = \lambda - 1$ is close to 0 as in Section 2. Then, the optimal tax τ_1^c satisfies

$$0 = \left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1 m_1\right) + \left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1\right) \varepsilon_{11}^c,$$

which we can rewrite as

$$\frac{\Lambda}{\psi_1} = \frac{\tau_1^p + m_1\tau_1^c}{p_1} \frac{m_1 + \varepsilon_{11}^c}{1 + \varepsilon_{11}^c}.$$

As long as $m_1 < 1$, the higher is the supply elasticity ε_S, the more the burden of the tax is shifted to the consumer, the higher is $\varepsilon_{11}^c < 0$, and the lower is the optimal tax.[53]

7.1.2 Productive Inefficiency at the Optimum

A canonical result in public finance (Diamond and Mirrlees 1971) shows that there is production efficiency at the optimum if there is a complete set of commodity taxes and either constant returns or fully taxed profits. We show that this result can fail even when the planner has a full set of commodity taxes of the $\boldsymbol{\tau}^c$ type (which would be enough to guaranty production efficiency in the traditional model), as long as there is not a full set of commodity taxes of the $\boldsymbol{\tau}^p$ type.

We start by considering the case where there is a full set of commodity taxes of the $\boldsymbol{\tau}^p$ type and show that production efficiency holds under some extra conditions. We denote by $\boldsymbol{q} = \boldsymbol{p} + \boldsymbol{\tau}^p$. We follow Diamond and Mirrlees (1971) and establish that production efficiency holds by assuming that the planner can control production, showing that the planner chooses an optimum on the frontier of the production possibility set. The corresponding planning problem is $\max_{\boldsymbol{q},\boldsymbol{\tau}^c} W\left(\left(v^h\left(\boldsymbol{q},\boldsymbol{\tau}^c,w\right)\right)_{h=1...H}\right)$ subject to the resource constraint $F\left(\boldsymbol{g} + \sum_h \boldsymbol{c}^h\left(\boldsymbol{q},\boldsymbol{\tau}^c,w\right)\right) \leq 0$.

Proposition 7.2 *With a full set of commodity taxes $\boldsymbol{\tau}^p$, production efficiency holds if either: (i) there are lump sum taxes and for all $\boldsymbol{q},\boldsymbol{\tau}^c$ and w, $v_w^h\left(\boldsymbol{q},\boldsymbol{\tau}^c,w\right) \geq 0$ for all h with a strict inequality for some h ; or (ii) for all $\boldsymbol{q}$, $\boldsymbol{\tau}^c$and w, there exists a good i with $v_{\boldsymbol{q}_i}^h\left(\boldsymbol{q},\boldsymbol{\tau}^c,w\right) \leq 0$ for all h with a strict inequality for some h.*

[53] Another way to see this is as follows. Consider the optimal tax with infinitely elastic supply $\varepsilon_S = \infty$ (a constant price p_1). It satisfies $\left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1 m_1\right) = 0$. Now imagine that $\varepsilon_S < \infty$. Then at this tax $\left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1\right) < 0$ so that $\left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1\right) \varepsilon_{11}^c > 0$ and by implication $\left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1 m_1\right) + \left(\Lambda c_1 - \tau_1^s \frac{\psi_1}{p_1} c_1\right) \varepsilon_{11}^c > 0$ This implies that increasing the tax improves welfare.

The proof is almost identical to the original proof of Diamond and Mirrlees (1971). Note however that the conditions $v_w^h(\boldsymbol{q}, \boldsymbol{\tau}^c, w) > 0$ or $v_{\boldsymbol{q}}^h(\boldsymbol{q}, \boldsymbol{\tau}^c, w) < 0$ can more easily be violated than in the traditional model without behavioral biases. Indeed, when agents misoptimize, it is entirely possible that the marginal utility of income be negative $v_w^h(\boldsymbol{q}, \boldsymbol{\tau}^c, w) < 0$. Loosely speaking, this happens if mistakes get worse as income increases. Similalry, it is entirely possible that $v_{q_i}^h(\boldsymbol{q}, \boldsymbol{\tau}^c, w) > 0$ for all i, since Roy's identity does not hold ($\frac{v_{q_i}}{v_w} \neq -c_i$). Failures of production efficiency could then arise even with a full set of commodity taxes $\boldsymbol{\tau}^p$. In the interest of space, we do not explore these conditions any further.

We now show that production efficiency can fail with a restricted set of commodity taxes $\boldsymbol{\tau}^p$, even if there is a full set of commodity taxes $\boldsymbol{\tau}^c$. Consider the following example. There are two consumption goods, 0 and 1, two types of labor, a and b, a representative agent with decision utility $u^s(c_0, c_1, l_a, l_b) = c_0 + U^s(c_1) - l_a - l_b$, and experienced utility $u^s(c_0, c_1, l_a, l_b) = u(c_0, c_1, l_a, l_b) - \xi_* c_1$, where $\xi_* > 0$ indicates an internality. For instance, c_1 could be cigarette consumption. Hence, the government would like to discourage consumption of good 1.

The production function for good i is $y_i = \left(\frac{l_{ia}}{\alpha_i}\right)^{\alpha_i} \left(\frac{l_{ib}}{1-\alpha_i}\right)^{1-\alpha_i}$, with $\alpha_i \in (0,1)$. As before, 0 is the untaxed good, $\tau_0 = 0$. The government can set taxes τ_1, τ_a and τ_b on good 1, labor of type a and labor of type b, and tax the employment of type a labor in sector 1. We assume that the consumer perfectly perceives taxes τ_a, τ_b, and prices p_0, p_1, p_a, p_b (the latter being the price of labor of type a, b). In addition, the government can set a tax τ_{1a} for the use of input a in the production of good 1. Note that production efficiency is equivalent to $\tau_{1a} = 0$.

Proposition 7.3 *If the consumer is fully inattentive to the tax τ_1, then the optimal tax system features production inefficiency: $\tau_{1a} > 0$. If the consumer is fully attentive to the tax τ_1, then the optimal tax system features production efficiency: $\tau_{1a} = 0$.*

The essence is the following—the government would like to lower consumption of good 1, which has a negative internality. However, agents do not pay attention to the tax τ_1 on good 1, therefore a tax on good 1 will not be effective. We assume that the government cannot use producer taxes. Hence, the government uses a tax $\tau_{1a} > 0$ on the input use in the production of good 1 (lowering production efficiency) to discourage the production of good 1, increase its price and discourage its consumption.

7.2 Atkinson-Stiglitz (1972)

Atkinson and Stiglitz (1972) show uniform commodity taxation is optimal if preferences have the form $u^h(c_0, \phi(\boldsymbol{C}))$, with $\boldsymbol{C} = (c_1, ..., c_n)$, ϕ homogeneous of degree 1, and c_0 (the untaxed good) might be leisure. We now investigate how to generalize this result with behavioral agents.

Proposition 7.4 *Consider the decision vs. experienced utility model. Assume that decision utility is of the form $u^{s,h}(c_0, \phi^s(\boldsymbol{C}))$ and that experienced utility is of the form $u^h(c_0, \phi(\boldsymbol{C}))$ with ϕ^s and*

ϕ homogeneous of degree 1. Then, if $\phi^s = \phi$, then uniform ad valorem commodity taxes are optimal (even though decision and experienced utility represent different preference orderings), but, if $\phi^s \neq \phi$, then uniform ad valorem commodity taxes are not optimal in general.

The bottom line is that with behavioral biases, it is no longer sufficient to establish empirically that expenditure elasticities for $(c_1, ..., c_n)$ are unitary.

Another relevant consideration has to do with time horizons. Consider a tax reform and assume away any link between periods for simplicity (say because agents do not have access to asset markets). Imagine a situation where, in the long-run, choices can be represented by a decision utility $u^{s,h}(c_0, \phi^s(\boldsymbol{C}))$, and welfare can be evaluated with an experienced utility $u^h(c_0, \phi(\boldsymbol{C}))$ with $\phi = \phi^s$. But, in the short-run as the tax code changes, agents misperceive taxes and, hence, make different choices. Then optimal time-varying taxes might be uniform in the long run but not in the short run. Likewise, if agents pay differential attention to taxes (at least in the short run), the Atkinson-Stiglitz (1972) neutrality result will fail.

8 Discussion

8.1 Novel Sufficient Statistics

Operationalizing our optimal tax formulas (17) and (42) requires taking a stand on the relevant sufficient statistics: social marginal value of public funds, social marginal utilities of income, elasticities, internalities, and externalities. For example, in the general Ramsey model, the optimal tax formula features the social marginal value of public funds λ, the social marginal utilities of income γ^h, consumption vectors $\boldsymbol{c}^h$, Slutsky matrices $\boldsymbol{S}^{C,h}$, and misoptimization wedges $\tilde{\boldsymbol{\tau}}^{b,h}$. All these sufficient statistics are present in the optimal tax formula of the traditional model with no behavioral biases, with the exception of misoptimization wedges $\tilde{\boldsymbol{\tau}}^{b,h}$. They are routinely measured by empiricists. They can be estimated with rich enough data on observed choices. The fact that agents are behavioral influences their value but does not change the way they should be estimated, but does require additional care. Indeed, with behavioral biases, these sufficient statistics might be highly context dependent, taking different values depending on factors that would be irrelevant in the traditional model, such as: the salience of taxes; the way taxes are collected; the complexity of the tax system; information about the tax system; the amount of time the tax system has been in place (allowing agents to become familiar with it); the presence of nudges, etc.

The misoptimization wedges $\tilde{\boldsymbol{\tau}}^{b,h}$, which summarize the effects of behavioral biases at the margin are arguably harder to measure. This poses a problem similar to the more traditional problem of estimating marginal externalities $\boldsymbol{\tau}^{\xi,h}$ to calibrate corrective Pigouvian taxes in the traditional model with no behavioral biases. The common challenge is that these statistics are not easily recoverable from observations of private choices. In both cases, it is possible to use a structural model, but more reduced-form approaches are also feasible in the case of behavioral biases.

Indeed, a common strategy involves comparing choices in environments where behavioral biases are attenuated and environments resembling those of the tax system under consideration. Choices in environments where behavioral biases are attenuated can be thought of as rational, allowing the recovery of experienced utility u^h as a utility representation of these choices, with associated indirect utility function v^h.[54,55] Differences in choices in environments where behavioral biases are present would then allow to measure the marginal internalities $\boldsymbol{\tau}^{b,h} = \boldsymbol{q} - \frac{u^h_{\mathbf{c}}}{v^h_w}$. For example, if the biases arise from the misperception of taxes so that $\boldsymbol{\tau}^{b,h} = \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h}$, then perceived taxes $\boldsymbol{\tau}^{s,h}$ could be estimated by comparing the environment under consideration to an environment where taxes are very salient or more generally where the environment is clearly understood by agents. If the biases arise because of temptation, then standard choices would reveal decision utility $u^{s,h}$. To the extent that agents are sophisticated and understand that they are subject to these biases, experienced utility u^h could be recovered by confronting agents with the possibility of restricting their later choice sets. In the terminology of Bernheim and Rangel (2009), this strategy uses refinements to uncover true preferences.

Another strategy (see e.g. Chetty, Kroft and Looney (2009) and Allcott, Mullainathan and Taubinsky (2012)) is to establish and isolate behavioral biases to find violations of the conditions imposed by rational choice (for example, showing that a demand elasticity depends on the salience of the tax). Yet another strategy, if behavioral biases arise from misperceptions, is to use surveys to directly elicit perceived taxes $\boldsymbol{\tau}^{s,h}$ (see e.g. Liebman and Zeckhauser (2004) and Slemrod (2006)).

Similarly, in the Mirrlees model, our optimal tax formula (42) features the distribution of income $h(z)$ and $h^*(z)$, the social marginal utilities of income $\gamma(z)$, the elasticities $\zeta^c(z)$, and $\zeta^c_{Q_{z^*}}(z)$ as well as $\widetilde{\tau}^b(z^*)$. All these sufficient statistics are present in the optimal tax formula of the traditional model with no behavioral biases, with the exception of $\zeta^c_{Q_{z^*}}(z)$ and $\widetilde{\tau}^b(z)$. Even though measuring the behavioral cross-influence $\zeta^c_{Q_{z^*}}(z)$ (which measures how changing the marginal tax rate at z^* impacts the labor supply at z) is typically neglected (because $\zeta^c_{Q_{z^*}}(z) = 0$ in the traditional model with no behavioral biases), measuring the elasticity $\zeta^c_{Q_{z^*}}(z)$ poses no conceptual difficulty as it can be recovered from observation of private choices but of course measuring it constitutes an interesting empirical challenge.[56] As for $\widetilde{\tau}^b(z)$, the same strategies that were discussed for $\widetilde{\tau}^{b,h}$ in the context of the Ramsey model above can be employed in this Mirrlees context as well.

In the misperception model, another, complementary approach is possible by using surveys (see e.g. De Bartolome 1995, Liebman and Zeckhauser 2004 and the references therein) to directly measure perceptions (in our model, this is the matrix $M_{ij} = \frac{\partial q_i^s(\boldsymbol{q},w)}{\partial q_j}$).

Finally, we have argued that the conditions (on the richness of the set of tax instruments) for

[54] Choices are more likely to reveal true preferences if agents have a lot of time to decide, taxes and long run effects are salient, and information about costs and benefits is readily available, etc.

[55] Our theory is invariant to the different cardinalizations of true preferences, as long as welfare weights are properly renormalized. For example, $\boldsymbol{\tau}^{b,h} = \boldsymbol{q} - \frac{u^h_{\mathbf{c}}}{v^h_w}$ is independent of the choice of cardinalization.

[56] That might help bridge the gap between micro and macro elasticities, as people are both influenced by their "local-micro" tax $(1 - q(z))$ and the "ambient" tax rates $(1 - q^d$, perhaps the average tax rate).

the supply elasticities not to enter direction in the optimal tax formulas are more stringent when agents are behavioral than in the traditional model. In some sense, supply elasticities and their empirical measurement are therefore more central to the behavioral approach than to the traditional approach.

8.2 Discussion of our Approach and Directions for Future Work

We now discuss a few limitations and potential extensions of our approach, some of which we plan to investigate in future work.

In our model, agents make mistakes, which the government may be able identify.[57] This approach, which is common but not uncontroversial, departs from the revealed preferences welfare paradigm and has elements of paternalism: the government tries to respect the agents' "true" preferences but recognizes that agents sometimes do not act in their own best interest (see Bernheim and Rangel 2009 for an in-depth discussion of this approach).

In addition, in our model, governments are benevolent and seek to rectify agents' mistakes through taxes and nudges. This normative benchmark leaves aside potentially important positive considerations. In practice governments also make mistakes, face various forms of political economy and institutional constraints, and may not be benevolent.

Moreover, despite our model's generality, there are categories of behavioral biases that it does not accommodate. First, our model only allows for intrapersonal but not for interpersonal behavioral deviations from the traditional model. For example, it leaves aside issues of fairness, relative comparisons, social norms, and social learning. Second, it is not ideally suited to capture information based behavioral phenomena, such as self and social signaling as a motivation for behavior, or the potential signaling effects of taxes and nudges (see e.g. Bénabou and Tirole 2006 and references therein).

9 Conclusion

We have generalized the main results of the traditional theory of optimal taxation to allow for large class of behavioral biases. Natural extensions would be to consider behavioral biases that cannot be captured by our model, such as interpersonal behavioral biases, or to move away from optimal taxation by analyzing optimal contracts. We plan to develop these issues in future work.

[57] Arguably, agents' mistakes can be persistent. For example, Slemrod (2006) argues that Americans overestimate on average the odds their inheritance will be taxed. Similarly, people seem to perceive average for marginal tax rates (Liebman and Zeckhauser 2004), and to overestimate the odds they'll move to a higher tax bracket (Benabou and Ok 2001). Second, our framework applies to situations where consumers do not maximize experienced utility. There, learning may be quite slow. For instance, people may persistently smoke too much, perhaps because of hyperbolic discounting (Laibson 1997).

References

Abaluck, Jason, and Jonathan Gruber, "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review* 101 (2011), 1180–1210.

Aguiar, Victor, and Roberto Serrano, "Slutsky Matrix Norms and the Size of Bounded Rationality," Working paper, Brown University, 2014.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky, "Energy Policy with Externalities and Internalities," *Journal of Public Economics*, 112 (2014), 72-88.

Allcott, Hunt, and Dmitry Taubinsky, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review* (forthcoming).

Allcott, Hunt, and Nathan Wozny, "Gasoline Prices, Fuel Economy, and the Energy Paradox," *Review of Economics and Statistics*, 96 (2014).

Anagol, Santosh, and Hoikwang Kim, "The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market," *American Economic Review*, 102 (2012), 576–93.

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

Atkinson, A.B. and J.E. Stiglitz, "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, 6 (1976), 55-75.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics* (2015).

Bénabou, Roland, and Efe Ok. "Social Mobility and the Demand for Redistribution: The POUM Hypothesis." *Quarterly Journal of Economics* (2001), 447-87.

Bénabou, Roland, and Jean Tirole. "Incentives and Prosocial Behavior," *American Economic Review*, 96 (2006), 1652-1678.

Bernheim, Douglas, and Antonio Rangel, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, 124 (2009), 51–104.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice", *Journal of Political Economy*, 121 (2013), 803–843.

Brown, Jennifer, Tanjim Hossain, and John Morgan, "Shrouded Attributes and Information Suppression: Evidence from the Field," *Quarterly Journal of Economics*, 125 (2010), 859–876.

Busse, Meghan, Christopher Knittel, and Florian Zettelmeyer, "Are Consumers Myopic? Evidence from New and Used Car Purchases," *American Economic Review*, 103 (2012), 220–256.

Caplin, Andrew, and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review* (forthcoming).

Carroll, Gabriel D., James Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Economics* 124 (2009), 1639-1674.

Chetty, Raj, "Behavioral Economics and Public Policy: A Pragmatic Perspective," *American*

Economic Review (forthcoming).

Chetty, Raj, "The Simple Economics of Salience and Taxation," NBER Working Paper, 2009.

Chetty, Raj, John Friedman, and Emmanuel Saez, "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review* 103 (2013), 2683–2721.

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

Chetty, Raj, and Emmanuel Saez, "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients." *American Economic Journal: Applied Economics*, 5 (2014), 1-31.

Cremer, Helmuth, and Pierre Pestieau. "Myopia, redistribution and pensions." *European Economic Review* 55 (2011), 165-175.

Davila, Eduardo, "Optimal Financial Transaction Taxes," Working Paper, 2014.

De Bartolome, Charles, "Which Tax Rate Do People Use: Average or Marginal?" *Journal of Public Economics* 56 (1995), 79-96.

DellaVigna, Stefano, "Psychology and Economics: Evidence from The Field," *Journal of Economic Literature*, 47 (2009), 315–372.

Diamond, Peter, "A Many-Person Ramsey Tax Rule," *Journal of Public Economics*, 4 (1975), 335-342.

Diamond, Peter, and James Mirrlees. "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review, 61* (1971), 8-27.

Ellison, Glenn and Sara Fisher Ellison, "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77 (2009), 427–452.

Finkelstein, Amy, "E-ZTAX: Tax Salience and Tax Rates," *Quarterly Journal of Economics*, 124 (2009), 969-1010.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics,* 129 (2014), 1661-1710.

Gabaix, Xavier, and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121 (2006), 505–540.

Gerritsen, Aart, "Optimal Taxation When People Do Not Maximize Well-being," Working Paper, University of Munich, 2014.

Glaeser, Edward, "Paternalism and Psychology," *University of Chicago Law Review,* 73 (2006), 133-156.

Glaeser, Edward, and Giacomo Ponzetto, "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions," *Journal of Public Economics,* 116 (2014), 89-105.

Goldin, Jacob, "Optimal Tax Salience," Working Paper, 2012.

Gruber, Jonathan and Botond Kőszegi, "Is Addiction 'Rational': Theory and Evidence," *Quarterly Journal of Economics*, 116 (2001), 1261-1303.

Handel, Ben, "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103 (2013), 2643-2682

Hastings, Justine, and Jesse Shapiro, "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks," *Quarterly Journal of Economics,* 128 (2013), 1449-1498.

Kaplow, Louis, "Myopia and the effects of social security and capital taxation on labor supply." *National Tax Journal*," 68 (2015), 7–32.

Kőszegi, Botond, and Adam Szeidl, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 128 (2013), 53–104.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics, 112* (1997), 443-477.

Liebman, Jeffrey B., and Richard J. Zeckhauser, "Schmeduling," Working Paper, 2004.

Lockwood, Ben, "Present Bias and the Optimal Taxation of Low Incomes," Working Paper, 2015.

Lockwood, Ben, and Dmitry Taubinsky, "Regressive Sin Taxes" Working Paper, 2015.

Loewenstein, George, and Ted O'Donoghue, "'We can do this the easy way or the hard way': Negative emotions, self-regulation, and the law," *University of Chicago Law Review, 73* (2006), 183-206.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. "Poverty impedes cognitive function," *Science*, 341 (2013), 976-980.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon, "A Reduced-Form Approach to Behavioral Public Finance," *Annual Review of Economics*, 4 (2012), 511-40.

O'Donoghue, Ted, and Matthew Rabin, "Optimal Sin Taxes," *Journal of Public Economics,* 90 (2006), 1825-1849.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118 (2003), 1-39.

Pigou, Arthur, *The Economics of Welfare*, 1920.

Ramsey Frank, "A Contribution to the Theory of Taxation," *Economic Journal*, 37 (1927), 47–61.

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

Saez, Emmanuel, "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, (117) 2002, 1039-1073.

Salanié, Bernard, *The Economics of Taxation*, MIT Press, 2011.

Sandmo, Agnar, "Optimal Taxation in the Presence of Externalities," *The Swedish Journal of Economics* (1975), 86-98.

Schwartzstein, Joshua, "Selective Attention and Learning," *Journal of the European Economic Association, 12* (2014), 1423–1452.

Sims, Christopher, "Implications of Rational Inattention," *Journal of Monetary Economics*, 50 (2003), 665–690.

Slemrod, Joel, "The Role of Misconceptions in Support for Regressive Tax Reform," *National Tax Journal* (2006), 57-75.

Spinnewijn, Johannes, "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs," *Journal of the European Economic Association, 13* (2014), 130-67.

Thaler, Richard H,"Mental Accounting and Consumer Choice," *Marketing Science,* 4 (1985), 199–214.

Thaler, Richard H., and Cass R. Sunstein, *Nudge,* Yale University Press, 2008.

Veldkamp, Laura, *Information Choice in Macroeconomics and Finance*, Princeton University Press, 2011.

Weitzman, Martin L., "Prices vs. Quantities," *The Review of Economic Studies*, 41 (1974), 477-491.

Woodford, Michael, "Inattentive Valuation and Reference-Dependent Choice," Working Paper, 2012.

10 Appendix: Notations

Vectors and matrices are represented by bold symbols (e.g. $\boldsymbol{c}$).

General notations

$\boldsymbol{c}$:consumption vector

h :index for household type h

L :government's objective function.

$\boldsymbol{m}, \boldsymbol{M}$:attention vector, matrix

$\boldsymbol{p}$: pre-tax price

$\boldsymbol{p}^s$: $s$ubjectively perceived price

$\boldsymbol{q} = \boldsymbol{p} + \boldsymbol{\tau}$: after-tax price

$\boldsymbol{q}^s$: $s$ubjectively after-tax perceived price

$\boldsymbol{S}_j, \boldsymbol{S}_j^C, \boldsymbol{S}_j^H$:Column of the Slutsky matrix when price j changes.

$u(\boldsymbol{c})$: true utility

$u^s(\boldsymbol{c})$: $s$ubjectively perceived utility

$v(\boldsymbol{p}, w)$: true indirect utility

w :personal income

W :social utility

$\lambda, \Lambda = \lambda - 1$:weight on revenue raised in planner's objective

ψ_i: demand elasticity for good i

$\boldsymbol{\tau}$: tax

$\boldsymbol{\tau}^s$: $s$ubjectively perceived tax

ξ: externality

Many-person Ramsey
$\mathcal{T}$: tax schedule (e.g. $T = \{\boldsymbol{\tau}, r_0\}$, where r_0 is a lump-sum transfer).
$\mathcal{Q}$: number of quantiles agents pay attention to
γ^h (resp. $\gamma^{\xi,h}$): marginal social utility of income (resp. adjusted for externalities)
θ :parametrization of perception functions
$\boldsymbol{\tau}^b$: misoptimization wedge
χ : nudge parameter

Nonlinear income tax
$g(z)$: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
$\boldsymbol{Q} = (q(z))_{z\geq 0}$: 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
$\zeta^c_{Q_{z*}}(z)$: compensated elasticity of earnings when the tax rate at z^* changes.
ζ^u : uncompensated elasticity of earnings

11 Appendix: Behavioral Consumer Price Theory

Here we develop behavioral consumer price theory with nonlinear budget. This nonlinear budget is useful both for conceptual clarity, and for the study of Mirrleesian nonlinear taxation. The agent faces a problem: $\max_{\boldsymbol{c}} u(\boldsymbol{c})$ s.t. $B(\boldsymbol{c}, \boldsymbol{p}) \leq w$. When the budget constraint is linear, $B(\boldsymbol{c}, \boldsymbol{p}) = \boldsymbol{p} \cdot \boldsymbol{c}$, so that $B_{p_j} = c_j, B_{c_j} = p_j$.

The agent, whose utility is $u(\boldsymbol{c})$, may not completely maximize. Instead, his policy is described by $\boldsymbol{c}(\boldsymbol{p}, w)$, which exhausts his budget $B(\boldsymbol{c}(\boldsymbol{p}, w), \boldsymbol{p}) = w$. Though this puts very little structure on the problem, some basic relations can be derived, as follows.

11.1 Abstract general framework

The indirect utility is defined as $v(\boldsymbol{p}, w) = u(\boldsymbol{c}(\boldsymbol{p}, w))$ and the expenditure function as $e(\boldsymbol{p}, \widehat{u}) = \min_{\boldsymbol{c}} B(\boldsymbol{c}, \boldsymbol{p})$ s.t. $u(\boldsymbol{c}, \boldsymbol{p}) \geq \widehat{u}$. This implies $v(\boldsymbol{p}, e(\boldsymbol{p}, \widehat{u})) = \widehat{u}$ (with $\widehat{u}$ a real number). Differentiating with respect to p_j, this implies

$$\frac{v_{p_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} = -e_{p_j}. \tag{49}$$

We define the compensated-demand based Slutsky matrix as:

$$\boldsymbol{S}_j^C(\boldsymbol{p}, w) = \boldsymbol{c}_{p_j}(\boldsymbol{p}, w) + \boldsymbol{c}_w(\boldsymbol{p}, w) B_{p_j}(\boldsymbol{c}, \boldsymbol{p})_{|\boldsymbol{c}=\boldsymbol{c}(\boldsymbol{p}, w)} \tag{50}$$

The Hicksian demand is: $\boldsymbol{h}(\boldsymbol{p}, \widehat{u}) = \boldsymbol{c}(\boldsymbol{p}, e(\boldsymbol{p}, \widehat{u}))$, and the Hicksian-demand based Slutsky matrix is defined as: $\boldsymbol{S}_j^H(\boldsymbol{p}, \widehat{u}) = \boldsymbol{h}_{p_j}(\boldsymbol{p}, \widehat{u})$.

The Slutsky matrices represent how the demand changes when prices change by a small amount, and the budget is compensated to make the previous basket available, or to make the previous utility available: $\boldsymbol{S}^C(\boldsymbol{p}, w) = \partial_{\boldsymbol{x}} \boldsymbol{c}(\boldsymbol{p} + \boldsymbol{x}, B(\boldsymbol{c}(\boldsymbol{p}, w), \boldsymbol{p} + \boldsymbol{x}))_{|\boldsymbol{x}=0}$ and $\boldsymbol{S}^H(\boldsymbol{p}, w) = \partial_{\boldsymbol{x}} \boldsymbol{c}(\boldsymbol{p} + \boldsymbol{x}, e(\boldsymbol{p} + \boldsymbol{x}, v(\boldsymbol{p}, w)))_{|\boldsymbol{x}=0}$ i.e., using (49),

$$\boldsymbol{S}_j^H(\boldsymbol{p}, w) = \boldsymbol{c}_{p_j}(\boldsymbol{p}, w) - \boldsymbol{c}_w(\boldsymbol{p}, w) \frac{v_{p_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} \tag{51}$$

In the traditional model, $\boldsymbol{S}^C = \boldsymbol{S}^H$, but we shall see that this won't be the case in general. [58]

We have the following elementary facts (with $\boldsymbol{c}(\boldsymbol{p}, w), v(\boldsymbol{p}, w)$ unless otherwise noted).

$$B_{\boldsymbol{c}} \cdot \boldsymbol{c}_w = 1, \qquad B_{\boldsymbol{c}} \cdot \boldsymbol{c}_{p_i} = -B_{p_i}, \qquad u_{\boldsymbol{c}} \boldsymbol{c}_w = v_w \tag{52}$$

The first two come from differentiating $B(\boldsymbol{c}(\boldsymbol{p}, w), \boldsymbol{p}) = w$. The third one comes from differentiating $v(\boldsymbol{p}, w) = u(\boldsymbol{c}(\boldsymbol{p}, w))$ with respect to w.

Proposition 11.1 (Behavioral Roy's identity) *We have*

$$\frac{v_{p_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} = -B_{p_j}(\boldsymbol{c}(\boldsymbol{p}, w), \boldsymbol{p}) + D_j(\boldsymbol{p}, w) \tag{53}$$

where

$$D_j(\boldsymbol{p}, w) = -\boldsymbol{\tau}^b(\boldsymbol{p}, w) \cdot \boldsymbol{c}_{p_j}(\boldsymbol{p}, w) = -\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H = -\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^C \tag{54}$$

and the misoptimization wedge is defined to be

$$\boldsymbol{\tau}^b(\boldsymbol{p}, w) = B_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w), \boldsymbol{p}) - \frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(\boldsymbol{p}, w)} \tag{55}$$

When the agent is the traditional rational agent, $\boldsymbol{\tau}^b = 0$. In general, $\boldsymbol{\tau}^b \cdot \boldsymbol{c}_w(\boldsymbol{p}, w) = 0$.

[58] See Aguiar and Serrano (2015) for a recent study of Slutsky matrices with behavioral models.

Proof: Relations (52) imply: $\boldsymbol{\tau}_b \cdot \boldsymbol{c}_w = \left(B_{\boldsymbol{c}} - \frac{u_{\boldsymbol{c}}}{v_w}\right) \boldsymbol{c}_w = 1 - 1 = 0$. Next, we differentiate $v(\boldsymbol{p}, w) = u(\boldsymbol{c}(\boldsymbol{p}, w))$

$$\begin{aligned}
\frac{v_{p_i}}{v_w} &= \frac{u_{\boldsymbol{c}} \boldsymbol{c}_{p_i}}{v_w} = \frac{\left(u_{\boldsymbol{c}} - v_w B_{\boldsymbol{c}} + v_w B_{\boldsymbol{c}}\right) \boldsymbol{c}_{p_i}}{v_w} \\
&= \frac{\left(u_{\boldsymbol{c}} - v_w B_{\boldsymbol{c}}\right) \boldsymbol{c}_{p_i}}{v_w} - B_{p_i} \text{ as } B_{\boldsymbol{c}} \cdot \boldsymbol{c}_{p_i} = -B_{p_i} \text{ from (52)} \\
&= -\boldsymbol{\tau}^b \cdot \boldsymbol{c}_{p_i} - B_{p_i} \qquad (56)
\end{aligned}$$

Next,

$$\begin{aligned}
D_j &= -\boldsymbol{\tau}^b \cdot \boldsymbol{c}_{p_j} = -\boldsymbol{\tau}^b \cdot \left(\boldsymbol{S}_j^H + \boldsymbol{c}_w(\boldsymbol{p}, w) \frac{v_{p_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)}\right) \text{ by (51)} \\
&= -\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H \text{ as } \boldsymbol{\tau}^b \cdot \boldsymbol{c}_w = 0 \qquad (57)
\end{aligned}$$

Likewise, (50) gives, using again $\boldsymbol{\tau}^b \cdot \boldsymbol{c}_w = 0$

$$D_j = -\boldsymbol{\tau}^b \cdot \boldsymbol{c}_{p_j} = -\boldsymbol{\tau}^b \cdot \left(\boldsymbol{S}_j^C - \boldsymbol{c}_w B_{p_j}\right) = -\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^C$$

□

Proposition 11.2 *(Slutksy relation modified) With $\boldsymbol{c}(\boldsymbol{p}, w)$ we have*

$$\begin{aligned}
\boldsymbol{c}_{p_j}(\boldsymbol{p}, w) &= -\boldsymbol{c}_w B_{p_j} + \boldsymbol{S}_j^H + \boldsymbol{c}_w D_j = -\boldsymbol{c}_w B_{p_j} - \boldsymbol{c}_w \left(\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H\right) + \boldsymbol{S}_j^H \\
&= -\boldsymbol{c}_w B_{p_j} + \boldsymbol{S}_j^C
\end{aligned}$$

and

$$\boldsymbol{S}_j^C - \boldsymbol{S}_j^H = \boldsymbol{c}_w D_j = -\boldsymbol{c}_w \left(\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H\right) \qquad (58)$$

Proof.

$$\begin{aligned}
\boldsymbol{c}_{p_j} &= \boldsymbol{c}_w \frac{v_{p_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} + \boldsymbol{S}_j^H \text{ by (51)} \\
&= \boldsymbol{c}_w \left(-B_{p_j} + D_j\right) + \boldsymbol{S}_j^H \text{ by Proposition 11.1}
\end{aligned}$$

Also, (50) gives: $\boldsymbol{c}_{p_j} = -\boldsymbol{c}_w B_{p_j} + \boldsymbol{S}_j^C$.□

Lemma 11.1 *We have*

$$B_{\boldsymbol{c}} \cdot \boldsymbol{S}_j^C = 0, \qquad B_{\boldsymbol{c}} \cdot \boldsymbol{S}_j^H = -D_j. \qquad (59)$$

Proof Relations (52) imply $B_{\boldsymbol{c}} \cdot \boldsymbol{S}_j^C = B_{\boldsymbol{c}} \cdot \left(\boldsymbol{c}_{p_j} + \boldsymbol{c}_w B_{p_j}\right) = -B_{p_j} + B_{p_j} = 0$. Also, $B_{\boldsymbol{c}} \cdot \boldsymbol{S}_j^H = B_{\boldsymbol{c}} \cdot \left(\boldsymbol{S}_j^C - \boldsymbol{c}_w D_j\right) = -D_j$. □

11.2 Application in Specific Behavioral Models

11.2.1 Decision-utility model

In the decision-utility model there is an experience utility function $u(\boldsymbol{c})$, and a perceived utility function $u^s(\boldsymbol{c})$. Demand is $\boldsymbol{c}(\boldsymbol{p}, w) = \arg\max_{\boldsymbol{c}} u^s(\boldsymbol{p}, \boldsymbol{c})$ s.t. $B(\boldsymbol{p}, \boldsymbol{c}) \leq w$.

Consider another agent who is rational with utility u^s. We call $v^s(\boldsymbol{p}, w) = u^s(\boldsymbol{c}(\boldsymbol{p}, w))$ his utility. For that other, rational agent, call $\boldsymbol{S}^{s,r}(\boldsymbol{p}, w) = \boldsymbol{c}_{\boldsymbol{p}}(\boldsymbol{p}, w) + \boldsymbol{c}_w(\boldsymbol{p}, w)' \boldsymbol{c}$ his Slutsky matrix. Given the previous results, the following Propsition is immediate.

Proposition 11.3 *In the decision-utility model, $\boldsymbol{S}_j^C = \boldsymbol{S}_j^{s,r}$ is the Slutsky matrix of a rational agent with utility $u^s(\boldsymbol{c})$. The misoptimization wedge is:*

$$\boldsymbol{\tau}^b = \frac{u^s_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v^s_w(\boldsymbol{p}, w)} - \frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(\boldsymbol{p}, w)}.$$

11.2.2 Misperception model

To illustrate this framework, we take the misperception model (i.e., the sparse max agent). It comprises a perception function $\boldsymbol{p}^s(\boldsymbol{p}, w)$ (which itself can be endogenized, something we consider later). The demand satisfies:

$$\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}^r(\boldsymbol{p}^s(\boldsymbol{p}, w), v(\boldsymbol{p}, w))$$

where $\boldsymbol{h}^r(\boldsymbol{p}^s, u)$ is the Hicksian demand of a rational agent with perceived prices $\boldsymbol{p}^s(\boldsymbol{p}, w)$.

Proposition 11.4 *Take the misperception model. Then, with $\boldsymbol{S}^r(\boldsymbol{p}, w) = \boldsymbol{h}^r_{\boldsymbol{p}^s}(\boldsymbol{p}^s(\boldsymbol{p}, w), v(\boldsymbol{p}, w))$ the Slutsky matrix of the underlying rational agent, we have:*

$$\boldsymbol{S}_j^H(\boldsymbol{p}, w) = \boldsymbol{S}^r(\boldsymbol{p}, w) \cdot \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) \tag{60}$$

i.e. $S^H_{ij} = \sum_k S^r_{ik} \frac{\partial p^s_k(\boldsymbol{p}, w)}{\partial p_j}$, where $\frac{\partial p^s_k(\boldsymbol{p}, w)}{\partial p_j}$ is the matrix of perception impacts. Also

$$\boldsymbol{\tau}^b = B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}) - \frac{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)}{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s) \cdot \boldsymbol{c}_w(\boldsymbol{p}, w)} \tag{61}$$

Given $B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c}) \cdot \boldsymbol{S}_j^H = 0$, we have:

$$D_j = (B_{\boldsymbol{c}}(\boldsymbol{p}, \boldsymbol{c}) - B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c})) \cdot \boldsymbol{S}_j^H = B_{\boldsymbol{c}}(\boldsymbol{p}, \boldsymbol{c}) \cdot \boldsymbol{S}_j^H \tag{62}$$

so that

$$D_j = \overline{\boldsymbol{\tau}}^b \cdot \boldsymbol{S}_j^H \text{ with } \overline{\boldsymbol{\tau}}^b = B_{\boldsymbol{c}}(\boldsymbol{p}, \boldsymbol{c}) - B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c}) \tag{63}$$

This implies that in welfare formulas we can take $\boldsymbol{\tau}^b = B_{\boldsymbol{c}}(\boldsymbol{p}, \boldsymbol{c}) - B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c})$ rather than the more cumbersome $\boldsymbol{\tau}^b = B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}) - \frac{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)}{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s) \cdot \boldsymbol{c}_w}$.

Proof Given $\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}^r(\boldsymbol{p}^s(\boldsymbol{p}, w), v(\boldsymbol{p}, w))$, we have $\boldsymbol{c}_w = \boldsymbol{h}^r_u v_w$. Then,

$$\begin{aligned}
\boldsymbol{S}^H_j &= \boldsymbol{c}_{\boldsymbol{p}_j}(\boldsymbol{p}, w) - \boldsymbol{c}_w(\boldsymbol{p}, w) \frac{v_{\boldsymbol{p}_j}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} = \boldsymbol{h}^r_{\boldsymbol{p}^s} \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) + \boldsymbol{h}^r_u v_{p_j} - \boldsymbol{c}_w \frac{v_{p_j}}{v_w} \\
&= \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) + \boldsymbol{h}^r_u v_{p_j} - \boldsymbol{h}^r_u v_w \frac{v_{p_j}}{v_w} \text{ as } \boldsymbol{c}_w = \boldsymbol{h}^r_u v_w \\
&= \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w)
\end{aligned}$$

Next, observe that the demand satisfies $u_{\boldsymbol{c}}(\boldsymbol{p}, w) = \Lambda B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c})$ for some Lagrange multiplier Λ, and that $B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c}) \cdot \boldsymbol{S}^r = 0$ for a rational agent (see equation (59) applied to that agent). So, $B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c}) \cdot \boldsymbol{S}^H = 0$.

$$\begin{aligned}
-D_j(\boldsymbol{p}, w) = \boldsymbol{\tau}^b \cdot \boldsymbol{S}^H_j &= \left(B_{\boldsymbol{c}} - \frac{u_{\boldsymbol{c}}}{v_w} \right) \cdot \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) = \left(B_{\boldsymbol{c}} - \frac{\Lambda B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c})}{v_w(\boldsymbol{p}, w)} \right) \cdot \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) \\
&= B_{\boldsymbol{c}} \cdot \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) = (B_{\boldsymbol{c}} - B_{\boldsymbol{c}}(\boldsymbol{p}^s, \boldsymbol{c})) \cdot \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w)
\end{aligned}$$

Finally, we have $\frac{u_{\boldsymbol{c}}}{v_w} = \Lambda B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)$ for some scalar $\Lambda > 0$. Given (52) $\frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(v, w)} = \frac{u_{\boldsymbol{c}}}{u_{\boldsymbol{c}} \cdot \boldsymbol{c}_w} = \frac{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)}{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s) \cdot \boldsymbol{c}_w}$ (indeed, both are equal to $\frac{u_{\boldsymbol{c}}}{u_{\boldsymbol{c}} \cdot \boldsymbol{c}_w}$).
□

We note that $u_{\boldsymbol{c}} \cdot \boldsymbol{S}^H = 0$ in the (static) misperception model (this is because $u_{\boldsymbol{c}} = \Lambda B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)$ for some scalar Λ, and $B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s) \cdot \boldsymbol{S}^H = 0$ from Proposition 11.3). This is not true in the decision-utility model.