Optimal Taxation with Behavioral Agents

Emmanuel Farhi  Xavier Gabaix
Harvard, CEPR and NBER  NYU, CEPR and NBER

February 9, 2015
Very preliminary and incomplete. Please do not circulate.*

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 and internalities. We revisit the three pillars of optimal taxation: Ramsey (taxation of goods to raise revenues), Pigou (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. Under some conditions, the optimal tax system is simple, in the sense that all tax rates are equal. For instance, all goods are optimally taxed at the same rate (or a few different rates), and the optimal income tax features just one marginal tax rate (or a few different marginal tax rates). This contrasts with the traditional optimal tax results, that generically features complex tax systems that depend on the details of the environment. Finally, we explore the Diamond-Mirrlees productive efficiency result and find that it is more likely to fail with behavioral agents.

*efarhi@fas.harvard.edu, xgabaix@stern.nyu.edu. For research assistance we thank Deepal Basak for research assistance, and for helpful comments we thank seminar participants at BU, Columbia, NYU, the UCL conference on behavioral theory, and Hunt Allcott. Gabaix thanks the NSF (SES-1325181) for support.
1 Introduction

This paper develops a systematic theory of optimal taxation with behavioral agents. Our framework is general. It allows for arbitrary behavioral biases, structures of demand, externalities, and population heterogeneity, as well as for a wide range of 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 result take the form of optimal tax formulas, which generalize the canonical formulas derived by Diamond (1975), Sandmo (1975) and Saez (2003). 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, social marginal of public funds, compensated demand elasticities, marginal externalities and equilibrium demands. Behavioral sufficient statistics are marginal internalities that arise only in behavioral models. They can be thought of as the ideal corrective taxes that would undo the underlying behavioral biases, such as misperception of taxes and internalities. 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); a modified optimal Pigouvian tax rule (one dollar of externality should be taxed by more than a dollar if people are inattentive to the tax); 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); 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.

Under some conditions, the optimal tax system is simple, in the sense that all tax rates are equal. For instance, all goods are optimally taxed at the same rate (or a few different rates), and the optimal income tax features just one marginal tax rate (or a few different marginal tax rates). This contrasts with the traditional optimal tax results, that generically features complex tax systems that depend on the details of the environment. The reason is that when the tax system is too complex, agents make mistakes (and anchor too much on some salient taxes). This is an
extra cost absent from the traditional model. Then, is the mistake can be large enough, the optimal
tax systems features just one uniform tax rates across goods or income brackets. It appears that
this is the first time that a “simple tax system” arises in public finance because of agents’ bounded
rationality.

Finally we turn to 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 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, what is required is a more stringent condition, namely that
there be a full set of commodity taxes which agents perceive like prices (in addition perhaps to
other commodity taxes which could be perceived differently from prices).

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. 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 fats or cigarettes (O’Donoghue and Rabin 2006).

Mullainathan, Schwarzstein and Congdon (2012) offer a rich overview of behavioral public fi-
nance. In particular, they derive optimality conditions for linear taxes, in a framework with a
binary action and a single good. Allcott, Mullainathan and Wozny (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 mar-
ginal tax rate for the average tax rate. A recent, independent paper by Lockwood and Wozny (2015)
studies 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

---

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
Ellison and Ellison (2009).
a general framework that reflects previous analyses, including misperceptions and internalities. One model we rely on for many illustrations is that sparse agent of Gabaix (2014), which builds on the burgeoning literature on inattention (Bordalo, Genaiolli Shleifer 2013, Caplin and Dean (2013), Chetty, Kroft and Looney (2009), Gabaix and Laibson (2006), Kőszei and Szeidl 2013, Schwartzstein 2014, Sims 2003, Woodford 2012). This agent misperceives prices (in a way that can be endogenized) to economize on attention (hence the name “sparse”), respects the budget constraint, in a way that keeps 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.

Our result that simple tax systems can be optimal with boundedly rational agents is conceptually different from other results in public finance regarding simple taxes with rational agents. First, some results prescribes zero or uniform taxes at the rational optimum: Diamond and Mirrlees (1971)’s productive efficiency result yields zero taxation of intermediate inputs; Atkinson and Stiglitz (1972, 1976) show that optimal commodity taxes are uniform under strong assumption regarding separability and homogeneity of preferences. Second, some results show that simple tax systems can sometimes be approximately optimal: see Farhi and Werning (2013) and Stantcheva (2014) for such results in dynamic environments.

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 studies the optimal nonlinear income tax problem. Section 6 offers a discussion, in particular of objects that would be important to measure. Section 5 offers extensions, in particular to production inefficiency. Section 6 offers a discussion, in particular of objects that would be important to measure.

2 The Basic Ramsey and Pigou Problems with Behavioral Agents

We start with an elementary treatment of the basic Ramsey problem with behavioral agents. We then move to Pigovian taxation.

2.1 Setup

The consumer There is a mass one of identical consumers. The representative consumer has a utility function $u(c)$ and budget constraint $q \cdot c \leq w$, where $c = (c_0, ..., c_n)$ is the consumption
vector, \( \mathbf{q} \) is the price vector, and \( w \) is wealth. If he were rational he would solve:

\[
\max_{\mathbf{c}} u(\mathbf{c}) \quad \text{s.t.} \quad \mathbf{q} \cdot \mathbf{c} \leq w
\]

However, he may be behavioral in some way. In the general model we will reason with a general demand \( \mathbf{c}(\mathbf{q}, w) \), which does not reflect optimization, but still exhausts the budget, \( \mathbf{q} \cdot \mathbf{c}(\mathbf{q}, w) = w \). This flexible and general modelling strategy will allow us to capture range of behavioral traits, from temptation to misperception of prices. To fix ideas, in this section we imagine that the demand function comes from the sparse max developed in Gabaix (2014):

\[
\text{smax } u(\mathbf{c}) \quad \text{s.t.} \quad \mathbf{q} \cdot \mathbf{c} \leq w.
\]

Here demand depends on both true prices \( \mathbf{q} \) and perceived prices \( \mathbf{q}^\sigma \). The demand \( \mathbf{c}^\sigma(\mathbf{q}, \mathbf{q}^\sigma, w) \) has the following characterization:

\[
\mathbf{c}^\sigma(\mathbf{q}, \mathbf{q}^\sigma, w) = \mathbf{c}^\sigma(\mathbf{q}^\sigma, w')
\]

where \( w' \) solves \( \mathbf{q} \cdot \mathbf{c}^\sigma(\mathbf{q}^\sigma, w') = w \). That is, the demand of a behavioral agent perceiving prices \( \mathbf{q}^\sigma \) and with budget \( w \) is the demand \( \mathbf{c}^\sigma(\mathbf{q}^\sigma, w') \) of a rational agent facing prices \( \mathbf{q}^\sigma \) and a slightly different budget \( w' \); the value of \( w' \) is chosen to satisfy the true budget constraint, \( \mathbf{q} \cdot \mathbf{c}^\sigma(\mathbf{q}^\sigma, w) = w \).

With this formulation the usual “trade-off” intuition applies in the space of perceived prices: \( \frac{\mathbf{a}_{1}}{\mathbf{a}_{2}} = \frac{\mathbf{a}_{1}^\sigma}{\mathbf{a}_{2}^\sigma} \), the ratio of marginal utilities is the ratio of perceived prices. For instance, with quasilinear utility \( u(\mathbf{c}) = c_0 + U(c_1, ..., c_n) \) with \( q_0 = q_0^\sigma = 1 \), we have, for \( i > 0 \), \( c_i^\sigma(\mathbf{q}, \mathbf{q}^\sigma, w) = c_i^\sigma(\mathbf{q}^\sigma) \): the demand of behavioral agent is the demand of a rational agent with perceived prices \( \mathbf{q}^\sigma \).

A function \( \mathbf{q}^\sigma(\mathbf{q}, w) \) maps true price into perceived prices. Starting in the next section, we’ll consider a very general one. For concreteness, in this section, we will often parametrize perceived prices as:

\[
q_i^\sigma(\mathbf{q}, w) = (1 - m_i) q_i^d(\mathbf{q}) + m_i q_i
\]

where \( q_i^d(\mathbf{q}) \) is a default price (that we will discuss later), and \( m_i \in [0, 1] \) is an attention parameter. \( m_i = 1 \) corresponds to full attention / rationality, \( m_i = 0 \) to full inattention. In the general model we will allow attention \( m \) to be endogenous, but we will take it as exogenous in this elementary section.

Hence, the demand is

\[
\mathbf{c}(\mathbf{q}, w) = \mathbf{c}^\sigma(\mathbf{q}, \mathbf{q}^\sigma(\mathbf{q}, w), w)
\]

i.e. the demand depends on both perceived prices and true prices, and perceived prices do depend on true prices. This demand is an instantiation of the more general behavioral demand.

**The government’s problem**  The government needs raises revenues by set a vector of commodity \( \mathbf{\tau} = (\tau_0, ..., \tau_n) \), where \( \tau_i \) is the linear tax of on good \( i \). We assume that one good (good 0,
perhaps “leisure”) is untaxed: \( \tau_0 = 0 \) throughout the paper.\(^2\) In addition, the production function is linear (in this section), so that, normalizing \( p_0 = 1 \), prices are set by the technological constraint: \( p \cdot y \leq w \), where \( w \) is the value \( p \cdot e \) of the initial endowment. Hence, the government’s problem is:

\[
\max_{\tau} u(c) + \lambda \tau \cdot c \text{ s.t. } c = c(p + \tau, w)
\]

where \( \lambda > 0 \) is the marginal utility of public funds.\(^3\) The agent faces the after-tax price \( q = p + \tau \).

The sparse agent has a perception of the tax that we parametrize in this section as:

\[
\tau^s_i = (1 - m_i) \tau^d(\tau) + m_i \tau
\]

for a “default” tax \( \tau^d \), which might depend on the true tax system \( \tau \). Hence, \( q_i = p_i + \tau_i \) and \( q^d_i = p_i + \tau^d \), and \( q^s_i = p_i + \tau^s_i \).

In this section, from now on, we consider a quasi-linear separable utility \( u(c) = c_0 + \sum_i u^i(c_i) \). As usual when considering the basic Ramsey problem, we work in the limit of small taxes (i.e., in the limit of \( \lambda \) close to 1). Without loss of generality we normalize production prices to 1 (\( p_i = 1 \)), so that \( \tau_i \) is the proportional tax on the good.\(^4\)

**Lemma 1** (Approximation of the government’s objective function). Calling \( L(\tau) = u(c) + \lambda \tau \cdot c \) the government’s objective function, we have \( L(\tau) - L(0) = L(\tau) + o(\|\tau\|^2) + O(\Lambda \|\tau\|^2) \), where

\[
L(\tau) = -\frac{1}{2} \sum_{i=1}^{n} \tau^s_i \psi_i y_i + \Lambda \sum_{i=1}^{n} \tau_i y_i
\]

where \( \Lambda = \lambda - 1 \), \( \tau^s_i \) is the perceived tax, and \( \psi_i \) the price-elasticity of demand for good \( i \) at zero taxes, and \( y_i \) is the expenditure when there is no tax.

We will use (4) as the government’s objective function in this section: this is main welfare term for \( \Lambda \) close to 0, which implies small taxes. It generalizes the usual formula with rational agents, \( L(\tau) = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{2} \tau^2_i \psi_i y_i + \Lambda \sum_{i=1}^{n} \tau_i y_i \).

The first term \( (\frac{1}{2} \sum_{i=1}^{n} \tau^2_i \psi_i y_i) \) in (4) captures the welfare distortions arising from taxation. Crucially it is the perceived tax \( \tau^s_i \), not the actual tax \( \tau_i \) that matters for welfare distortions. 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 \( \tau_i \), not the perceived tax, that matters.

\(^2\)Otherwise, the taxation problem with a representative agent is trivial as the government can replicate lump-sum taxes via a uniform tax on all goods.

\(^3\)If the government needs to raise revenues \( G \) from taxes, the problem is \( \max_{\tau} u(c) \text{ s.t. } \tau \cdot c \geq G \) and \( c = c(p + \tau, w) \). Then, \( \lambda \) is endogenous and equal to the Lagrange multiplier of the problem.

\(^4\)This lemma is anticipated in Chetty, Looney and Kroft (2009).
2.2 Optimal Taxes in the Basic Ramsey Problem

Let us first revisit the traditional Ramsey solution with rational agents. Then \( \tau_i^* = \tau_i \), and formula (4) gives: 

\[
L(\tau) = \frac{-1}{2} \sum_i \frac{1}{2} \tau_i^2 \psi_i y_i + \Lambda \sum_i \tau_i y_i.
\]

Forming \( \frac{\partial L}{\partial \tau_i} = -\tau_i \psi_i y_i + \Lambda y_i = 0 \), so we obtain the Ramsey’s inverse-elasticity rule \( \tau_i = \tau_i^R \) with

\[
\tau_i^R = \frac{\Lambda}{\psi_i}.
\]  

We now move to the behavioral agent. We take here the perceptions to be as in (3). It proves convenient to separate the role of \( \tau^d \) by writing

\[
L(\tau, \tau^d) = -\frac{1}{2} \sum_i \left( (1-m_i) \tau^d + m_i \tau_i \right)^2 \psi_i y_i + \Lambda \sum_i \tau_i y_i
\]

**Lemma 2** At the optimum in the basic Ramsey problem, the optimal tax satisfies: \( \frac{dL}{d\tau_i} = 0 \) with

\[
\frac{dL(\tau, \tau^d)}{d\tau_i} = -\tau_i^s \psi_i y_i + \Lambda y_i + \frac{\partial L(\tau, \tau^d)}{\partial \tau^d} \frac{\partial \tau^d(\tau)}{\partial \tau_i}
\]  

with \( \frac{\partial L}{\partial \tau^d} = -\sum_j \tau_j^s (1-m_j) \psi_j y_j \).

**Proof** We have

\[
\frac{\partial L(\tau, \tau^d)}{\partial \tau_i} = -\tau_i^s \psi_i y_i + \Lambda y_i
\]

\[
\frac{\partial L(\tau, \tau^d)}{\partial \tau^d} = -\sum_j \tau_j^s (1-m_j) \psi_j y_j
\]

The total impact of changing \( \tau_i \) is: \( \frac{dL}{d\tau_i} = \frac{\partial L}{\partial \tau_i} + \frac{\partial L}{\partial \tau^d} \frac{\partial \tau^d(\tau)}{\partial \tau_i} \).

We state the following important consequence from (6).

**Proposition 1** When \( \tau^d \) is fixed at 0 and \( m_i > 0 \), at the optimum, we have the following inverse-elasticity rule modified by inattention:

\[
\tau_i = \frac{\Lambda}{m_i^2 \psi_i}, \quad \tau_i^s = \frac{\Lambda}{m_i \psi_i}.
\]  

Optimal taxes are higher at \( \tau_i = \frac{\Lambda}{m_i^2 \psi_i} \) than they would be in the traditional Ramsey solution with rational agents at \( \tau_i = \frac{\Lambda}{\psi} \). Loosely speaking, this is because inattention makes agents less elastic: given partial attention \( m_i \leq 1 \), the effective elasticity is \( m_i \psi_i \) rather than the parametric elasticity \( \psi_i \). In the spirit of the traditional Ramsey formula, lower elasticity leads to higher optimal
taxes.\footnote{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 the correct \( \tau_i = \frac{\Lambda}{m_i^* \psi_i} \). To gain intuition for the correct formula, consider the effect of a marginal increase in \( \tau_i \). The marginal cost in terms of increased distortions is \( \tau^*_i m_i \psi_i y_i \) (equation (7)). At the optimum, the marginal cost and the marginal benefit are equalized. The result is that \( \tau^*_i = \frac{\Lambda}{m_i^* \psi_i} \), i.e. it is the perceived tax \( \tau^*_i \) that is inversely related to the effective elasticity \( m_i \psi_i \). This in turns implies \( \tau_i = \frac{\tau^*_i}{m_i} = \frac{\Lambda}{m_i^2 \psi_i} \).

When \( m_i > 0 \) and \( \tau^d \) depends on \( \tau \), equation (6) also gives
\[
\tau_i = \frac{\Lambda}{m_i^2 \psi_i} + \frac{\partial L}{\partial \tau^d} \frac{\partial r^d}{\partial \tau_i} - \frac{1}{m_i} \tau^d, \quad \tau^*_i = \frac{\Lambda}{m_i \psi_i} + \frac{\partial L}{\partial r^d} \frac{\partial \tau^d}{\partial \tau_i}
\]

Hence (as \( \frac{\partial L}{\partial \tau^d} \leq 0 \), the optimal tax on a good \( i \) is lower when the tax on this good increase the default tax (the term \( \frac{\partial \tau^d}{\partial \tau_i} \)). We next explore a consequence of this cross-perception effect \( \frac{\partial \tau^d}{\partial \tau_i} \).

2.3 Simple Tax Systems at the Optimum

2.3.1 Condition yielding a uniform tax rate

In this subsection we take a stand on the default tax \( \tau^d \), and suppose that it has the following form:
\[
\tau^d = \alpha \tau_{\text{max}} + \beta \tau^\omega,
\]
where \( \alpha > 0, \beta \geq 0, \tau_{\text{max}} = \max_i \tau_i \) and \( \tau^\omega = \sum_i \omega_i \tau_i \) is the average tax, where the weights \( \omega_i \) satisfy \( \omega_i \geq 0 \) and \( \sum_i \omega_i = 1 \).

This formulation is meant to capture the psychology of salience, as the maximum tax is one salient number in a tax system. This is a plausible conjecture, discussed further in Section 6.1. We note that the literature on robustness and min-max behavioral also puts a high weight on the worse-case dimension.

\textbf{Proposition 2} For a given \( \alpha > 0 \), the optimal tax system is simple, i.e. features the same tax for all goods, if \( \beta \) and the \( m_i \)'s are sufficiently small. In the case \( \alpha = 1, \beta = 0 \), we have the following explicit sufficient condition: for all \( i \), \( m_i \psi_i < \bar{\psi} \) with \( \bar{\psi} = \frac{\sum_i y_i \psi_i}{\sum_i y_i} \).

The condition \( m_i \psi_i \leq \bar{\psi} \) means that the attention \( m_i \) to the price of goods with high elasticities (those that should have a low tax rates) are low enough.\footnote{When \( \beta > 0, \alpha + \beta = 1 \), Proposition 44 in the online appendix gives a quantitative sufficient condition: we have}
To get an idea for the formal proof, consider the case $m_i = \beta = 0$. Then, if there is a good $i$ with $\tau_i < \tau_{\text{max}}$, equation (6) gives $\frac{dL(\tau)}{d\tau_i} = \Lambda y_i - \tau_i^* m_i \psi_i y_i + \frac{\partial L(\tau, \tau^d)}{\partial \tau_i} \frac{\partial \tau^d(\tau)}{\partial \tau_i} = \Lambda y_i > 0$, (as $m_i = 0$ and $\frac{\partial \tau^d(\tau)}{\partial \tau_i} = 0$): the planner wants to increase the tax on the good, as this does not impact any perceived tax. This holds up to the point where $\tau_i = \tau_{\text{max}}$. Hence, at the optimum, all taxes are equal to $\tau_{\text{max}}$.

### 2.3.2 Conditions yielding a small number of different tax rates

We now generalize this insight, and suppose that people use the following “quantile-based default model”:

$$
\tau^d(\tau) = \beta \tau_\omega + \sum_{k=1}^{Q} \alpha_k \tau^{(r_k)}
$$

where $r$ is the $r$–th quantile (with the convention that $\tau^{(1)} = \tau_{\text{max}}$, $\tau^{(0)} = \tau_{\text{min}}$), and the $r_k$’s correspond to different quantiles of taxes, and $\alpha_k$ is the weight on quantile $r_k$. We assume that people pay attention to the maximum tax, so that $r_1 = 1$ and $\alpha_1 > 0$. As before, $\tau_\omega$ is the mean tax with weight $\omega = (\omega_i)_{i=1...n}$. We assume $\beta, \alpha_k \geq 0$.

**Proposition 3** Suppose that people pay attention to $Q$ different quantiles, including a positive weight on $\tau_{\text{max}}$, as in the “quantile-based default model” (10). Provided that $m_i$ and $\beta$ are small enough, at the optimum there are at most $Q$ different tax rates.

**Proof** We start with the case $m = 0$, $\beta = 0$. We will show that all $\tau_i$ must be exactly at some quantile $\tau^{(r_k)}$. Indeed, suppose towards a contradiction that $\tau_i$ is not equal to any of the quantiles $\tau^{(r_k)}$, we have $\frac{\partial \tau^d}{\partial \tau_i} = 0$, and equation (6) gives: $\frac{dL}{d\tau_i} = \Lambda y_i - \tau_i^* m_i \psi_i y_i + \beta \omega_i \frac{\partial L}{\partial \tau_i} = \Lambda y_i > 0$. Hence, the planners wants to increase $\tau_i$ locally: the value of $\tau_i$ was not optimal. We have reached the desired contradiction.

In the general case with $m, \beta$ close to 0, the proof is the same, by continuity. We provide a rigorous proof in Proposition 16, which is a strict generalization of Proposition 3. □

One could also imagine that consumers form a different default tax for each category of goods. For instance, suppose that the default tax for all food items puts a high weight on the maximum tax within all food items, and attention is low (as in Proposition 2), there will be a uniform tax for all food items at the optimum.

---

simple taxes when $0 \leq -m_i \frac{\omega_i}{\overline{\omega}} s_i + s_i - (1 - \overline{m}) \beta \omega_i \leq \alpha (1 - \overline{m})$, for $\overline{m} = \frac{\sum_i m_i \psi_i y_i}{\sum_i \psi_i y_i}$ and $s_i = \frac{y_i}{\sum_j y_j}$. This means that the $m_i$ are low enough, $\frac{\omega_i}{\overline{\omega}}$ high enough, and $\frac{\beta \omega_i}{s_i}$ is low enough. Indeed, take the case where all $m_i = 0$: then the condition states: $0 \leq s_i - \beta \omega_i \leq \alpha$, hence we must have $\beta \omega_i \leq s_i$, i.e. the weight $\omega_i$ on the $\tau^d$ cannot be too much higher than the true weight $s_i$ in the basket. Otherwise, the planner would be tempted to set a very low, negative tax on a good $i$, a lower enough the perceived tax.

---

9
2.4 Correcting Externalities and Internalities: The Basic Pigou Problem Revisited

2.4.1 Corrective tax for inattentive agents

We remain in the quasilinear case, and consider one good. The representative agent maximizes 
\[ u(c_0, c) = c_0 + U(c) \] subject to \[ c_0 + pc \leq w. \] Here \( c \) stands from the consumption the good 1 (we could call it \( c_1 \), but expressions are cleaner by calling it \( c \)). Hence, he solves:

\[
\max_c U(c) - pc. \tag{11}
\]

However, there is a negative externality that depends on the aggregate consumption of good 1 (e.g., second-hand smoke), so that total utility is \( c_0 + U(c) - \xi c \). Alternatively, \( \xi \) 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 exactly identical in both cases.

To focus on the corrective role of taxes, we assume \( \Lambda = 0 \) and that the government can rebate tax revenues lump-sum to consumers. The government’s objective problem is as follows.

\[
\max_c U(c) - (p + \xi) c. \tag{12}
\]

To attempt to correct the externality/internality, the government can set a tax \( \tau \).

If the agent is rational, his problem becomes \( \max_c U(c) - (p + \tau) c \). The optimal tax is \( \tau = \xi \), as the agent maximizes the planner’s objective. This is the classic Pigouvian tax that makes the agent exactly internalize the externality/internality.

Now consider a behavioral agent, who perceives only a fraction \( m \) of the tax. Then, his decision problem is:

\[
\max_c U(c) - (p + m\tau) c. \tag{13}
\]

So that the agent’s problem mimics that of the planner (equation (12)), the tax must be \( \tau = \frac{\xi}{m} \).

We record this simple result.7

Proposition 4 If the consumer has attention \( m \) to the tax and the externality/internality is \( \xi \), the optimal Pigovian tax is \( \tau = \frac{\xi}{m} \). The first best is then achieved.

Suppose for concreteness that a good has a negative externality of $1. With rational agents, it should be taxed by $1. However, suppose that all agents perceive only half the tax. Then, the good should be taxed by $2 – as the first best is achieved when all agents perceived a tax of $1.

---

7 The fact that the tax should be higher when it is underperceived was qualitatively anticipated by Mullainathan, Schwartzstein and Cogdon (2012).
If different consumers have different $m$, Proposition 4 suggests that no uniform tax will correct all of them. Hence, heterogeneity in attention will prevent taxes from achieving the first best. We next explore this more thoroughly.

2.4.2 Accounting for heterogeneity

In the previous analysis, there was just one homogenous type of consumers. We next analyze what happens with several types, indexed by $h = 1 \ldots H$. Agent of type $h$ maximizes $u^h (c_0^h, c_h) = c_0^h + U^h (c_h)$, and the associated externality/ internality is $\xi^h c_h$. He pays an attention $m_h$ to the tax, i.e. perceives $\tau_h^s = m_h \tau$ (recall that we assume $\tau^d = 0$). The planner’s objective is utilitarian, i.e. he maximizes

$$\sum_{h} U^h (c_h) - (p + \xi^h) c_h.$$  \hspace{1cm} (14)

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

To make things transparent, we can use: $U^h (c) = \frac{a^h c - \frac{1}{2} c^2}{\Psi}$, which implies the demand: $U^h = \frac{a^h - \Psi}{\frac{\Psi}{a^h}} = q^s$, hence $c^h (q^s) = a^h - \Psi q^s$. The expressions in the rest of this section are exact with that quadratic utility function. For generality utility functions, they hold provided that they are understood as the leading order terms in a Taylor expansion around an economy with no heterogeneity.\(^8\)

The social loss compared to the first best is:

$$L^{\text{tax}} = -\frac{\Psi}{2} \sum_{h} (m^h \tau - \xi^h)^2.$$  \hspace{1cm} (15)

Hence, no tax will be above to achieve the first best unless all agents have the same $\xi^h/m^h$.

**Proposition 5** (Heterogeneity in attention or externality/internality creates welfare losses) The optimal tax is:

$$\tau^* = \frac{E [\xi_h m_h]}{E [m_h^2]}.$$  

The welfare losses from that optimal tax stemming from the heterogeneity in attention and/or externality/internality are:

$$L^{\text{tax}} = L^{\text{no \, tax}} \frac{E [\xi_h^2] E [m_h^2]}{E [m_h^2]} (E [\xi_h m_h])^2.$$  

where $L^{\text{no \, tax}} = -\frac{1}{2} \Psi H E [\xi_h^2]$ is the welfare loss without corrective taxation.

A key message of Proposition 5 is that heterogeneity in attention creates welfare losses. Take the case of a uniform externality/internality $\xi^h = \xi$, but heterogeneous attention to the tax. Then,

\(^8\)Our results extend readily to differences $\Psi_h$ in $\Psi$ across agents, where $\Psi^h = -1/U_{cc}^h (c^*_h)$.  

11
even an optimal tax cannot fully correct the externality/internality, unlike in the basic Pigovian model. The welfare loss is \( L^{\text{no tax}} = \frac{\text{var}(m_h)}{E[m_h]} \), so increase with the variance of \( m_h \).

### 2.4.3 Quantity vs Price regulation

**Quantity restriction** Suppose the government imposes a uniform quantity restriction, mandating \( c_h = c^* \).

**Proposition 6** The optimal quantity restriction \( c^* \) is: \( c^* = E[c^*_h] \). The social loss is then: \( L^Q = -\frac{1}{2} H \Psi \text{var}(c^*_h) \).

This proposition shows the key disadvantage of uniform quantity restrictions is that the optimal quantity across agents is typically variable (the \( \text{var}(c^*_h) \) term).

**Corollary 1** (Quantity vs Price regulation under inattention) **Quantity restrictions are better than taxation** \((L^Q \geq L^{\text{tax}}) \) if and only if:

\[
\frac{1}{\Psi} \text{var}(c^*_h) \leq \Psi E[\xi^2_h] E[m^2_h] - (E[\xi_h m_h])^2 \frac{1}{E[m^2_h]}
\]

Several insights can be gleaned from this formula, reminiscent of Weitzman (1974). First, quantity restrictions are better if attention \((m_h)\) is very heterogeneous. Taxes are better if there optimal consumption is very heterogeneous. Second, when the demand elasticity is low \((\text{low } \Psi)\), quantity restrictions are worse. When elasticity is low, agents suffer when they deviate from their optimal quantity, and preventing them from consuming the good is very costly. Third, for small enough externalities, taxes are better than quantity restrictions. Indeed, there are only second order losses from the externality \((E[\xi^2_h])\), while the quantity restriction discretely lowers welfare (to the zero-th order) even when the externality is very small.\(^9\)

### 2.4.4 Optimal nudges

Here we provide a minimalist and simplistic model of nudges. Section 3.4 will provide a more general model. A subset \( N \) of agents are nudgeable; their mass is \( H_N \). We posit that a nudge indexed by \( \chi \) persuades them to consume a quantity \( \chi \) of the good.

**Proposition 7** The optimal nudge is the average optimal quantity amongst nudgeable agents: \( \chi^* = E[c^*_h \mid h \in N] \). Compared to the first best, the social loss is:

\[
L^\chi = -\frac{1}{2} H_N \Psi \text{var}(c^*_h \mid h \in N) - \frac{1}{2} (H - H_N) \Psi E[\xi^2_h \mid h \notin N]
\]

\(^9\)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).
Proof The optimal nudge $\chi$ is chosen exactly like in Proposition 6. The losses from the unnudged agents is as in Proposition 5. □

The proposition highlights that the optimal nudge is the average optimal quantity for nudged agents, rather than for the total population.

**Corollary 2** (Nudges vs Price regulation under inattention) *Nudges are better than taxation* ($L^X \geq L^{tax}$) *if and only if:*

$$
\frac{H_N}{\Psi} \text{var} (c^*_h \mid h \in N) + (H - H_N) \Psi E \left[ \xi^2_h \mid h \notin N \right] \leq H \Psi E \left[ \xi^2_h \mid h \notin N \right] - (E [\xi_h m_h])^2
$$

(17)

*Nudges are better than quantity restrictions* ($L^X \geq L^Q$) *if and only if:*

$$
\frac{H_N}{\Psi} \text{var} (c^*_h \mid h \in N) + (H - H_N) \Psi E \left[ \xi^2_h \mid h \notin N \right] \leq \frac{H}{\Psi} \text{var} (c^*_h).
$$

(18)

Compared to taxes and quantity restrictions, nudge might allow to target a subset of particularly biased agents. Nudges are an attractive option when nudged agents are relatively homogeneous (low $\text{var} (c^*_h \mid h \in N)$) and unnudged agents require little intervention (low $E [\xi^2_h \mid h \notin N]$).

To illustrate the potential advantage of nudges, consider the following example. There are only internalities. A fraction $1 - f$ of agents are rational, and can’t be nudged, but also don’t have internalities: their $\xi^h$ is 0. A fraction $f$ of agents are nudgeable and their first best optimal consumption $c^*_h$ is $c^*$ independent of $h$. Then, the optimal nudge $\chi = c^*$ achieves the first best. By contrast in general taxes and quantity restrictions will not achieve the first best.

## 3 Optimal Commodity Taxation with Redistribution: General Approach

### 3.1 Basic setup

#### 3.1.1 The consumer

**Abstract, general version** We now consider a consumer with demand $c(q, w)$, that potentially incorporates behavioral biases (e.g. internalities, misperceptions, etc.), where $q$ is the price vector, $w$ is the budget. It still verified $q \cdot c(q, w) = w$. His true (“experienced”) utility is $u(c)$.

Note that this imposes that producer prices $p$ and taxes $\tau$ matter to the consumer only through $q = p + \tau$. In Sections 5.1 and 5.2, we relax this assumption and consider a demand function of the form $c(p, \tau, w)$. Note however that this distinction is immaterial if producer prices $p$ are fixed, which we assume in this section, but which we relax in Section 5.2.

The following concepts are useful. The indirect utility function is $v(q, w) = u(c(q, w))$. The expenditure function is $e(q, \tilde{u}) = \min_w w$ s.t. $v(q, w) \geq \tilde{u}$. The Hicksian demand is $h(q, \tilde{u}) = \ldots$
behavioral consumer theory” is in Appendix 9.1. We show here the highlights.

The Slutsky relation is modified as: \( \tau^b(q, w) \cdot S^H_j(q, w) \) is a discrepancy term in Roy’s identity that arises from a failure of the envelope theorem because agents do not fully optimize. With the traditional model, \( \tau^b(q, w) = 0 \).

The Slutsky relation is modified as:

\[
c_{q_j}(q, w) = -c_w(q, w)c_j(q, w) + S^H_j(q, w) - c_w(q, w)\tau^b(q, w) \cdot S^H_j(q, w)
\]  

(20)

Hence, the Slutsky relation itself is modified by the wedge \( \tau^b(q, w) \). One can also define a Slutsky matrix as: \( c_{q_j}(q, w) = -c_w(q, w)c_j(q, w) + S^C_j(q, w) \). This definition of \( S^C_j(q, w) \) corresponds to an income compensation for a change in the price \( q_j \), while the definition of \( S^H_j(q, w) \) corresponds to a compensation utility compensation. Only in the traditional model do the two concepts coincide. In general, \( S^C_j(q, w) = S^H_j(q, w) - c_w(q, w)\tau^b(q, w) \cdot S^H_j(q, w) \).

From here on, we omit the dependence of all functions on \( (q, w) \), unless an ambiguity arises.

Decision vs. experienced utility model The demand function could arise from a “decision utility” \( u^b \), i.e. \( c(q, w) = \max_c u^b(c) \) s.t. \( q \cdot c \leq w \). However the true “experienced” utility remains \( u(c) \). Then, \( \tau^b = \frac{\partial u^b}{\partial w} - \frac{\partial u}{\partial w} \) and \( S^C_j = S^H_j \) is the Slutsky matrix of an agent with preference \( u^b \).

For instance, if \( c = (c_0, c_1) \) and \( u^b(c) = c_0 + U^b(c_1) \) and \( u(c) = c_0 + U(c_1) \) where \( c_1 \) is the consumption of ice cream and \( c_0 \) an outside good, \( U^b(c_1) \) might be immediate pleasure from eating the ice cream, while \( U(c_1) \) represent those immediate pleasures minus future pain from the extra weight gained (see for instance O’Donoghue and Rabin 2006). Then \( \tau^b = (\tau^b_0, \tau^b_1) \) where \( \tau^b_0 = 0 \) and \( \tau^b_1 = U^b_c - U_c \) reflects the wedge between marginal decision and experienced utility. In the ice cream example, \( \tau^b_1 \) is positive.

Misperception model We revisit the model based on misperceptions outlined in Section 2. We again highlight results derived in the appendix. Given true prices \( q \), perceived prices \( q^\prime \), budget \( w \), and the demand \( c^*(q, q^*, w) = \max_{c \mid q^*} u(c) \) s.t. \( q \cdot c \leq w \) is the consumption vector \( c \).
satisfying \( u_c(c) = \lambda^* q^* \) for some \( \lambda^* > 0 \) and \( q \cdot c = w \). We are also given a perception function \( q^*(q, w) \). Then, demand is: \( c(q, w) = c^*(q, q^*(q, w), w) \). Then, the ideal corrective tax is formally \( \tau^b = q - \frac{q^*}{q^*-e_w} \), but in all formulas it can be taken to be:

\[
\tau^b = q - q^*.
\]

Indeed, we have \( q^* \cdot S^H = 0 \) in the misperception model. This fleshes out the idea that \( \tau^b \) represents the impact of the discrepancy between true prices and perceived prices.

Given this demand function \( c^*(q, w) \), the notions of the abstract model apply and receive a concrete interpretation. First, we define \( M_{ij}(q, w) = \frac{\partial q^*(q, w)}{\partial q_j} \) and the vector

\[
M_j(q, w) = \frac{\partial q^*(q, w)}{\partial q_j}
\]

which is the marginal impact of a change in true price \( q_j \) on the perceived price \( q^* \). We then define the Hicksian demand of a fictitious rational agent as: \( h^r(q^*, \hat{u}) = \arg\min_q \ q^* \cdot c \ \text{s.t.} \ u(c) \geq \hat{u} \). Then, we define \( S^H_j(q^*, \hat{u}) = h^r_q(q^*, \hat{u}) \) the “rational” Slutsky matrix for the rational agent. We also define \( S^r(q, w) = S^r(q^*(q, w), v(q, w)) \). Then, the Slutsky matrix \( S^H \) we saw in the abstract model has a concrete interpretation:

\[
S^H_j(q, w) = S^r(q, w) \cdot \frac{\partial q^*(q, w)}{\partial q_j}.
\]

i.e. \( S^H_{ij}(q, w) = \sum_k S^r_{ik}(q, w) \frac{\partial q^*(q, w)}{\partial q_j} \). This is, when the price of good \( j \) changes, it creates a change in the perceived prices \( q^*_k \), which in turn changes (via the rational Slutsky matrix \( S^r \)) the demand for good \( i \). The matrix \( S^H_{ij} \) captures this total effect.

**Isomorphism between the abstract and misperception models** We now link the abstract demand to the misperception framework section 2 started with. Given an abstract demand \( c(q, w) \), we can define a virtual perceived price as a shadow price \( q^* \)

\[
q^*(q, w) = \frac{u_c(c(q, w))}{v_w(q, w)}
\]

This defines a sparse max agent with this perception function. Lemma 7 in the Appendix shows that \( c(q, w) = c^*(q, q^*(q, w), w) \). This is, the demand function of a general abstract consumer can be represented as that of a sparse consumer with perceived prices \( q^*(q, w) \).

---

10 If there are several such \( \lambda \), we take the lowest one, which is also the utility-maximizing one.

11 For instance, in specification (2), \( M_{ij} = m_{ij} 1_{i=j} + (1 - m_{ij}) \frac{\partial q^*}{\partial q_j} \).

12 Any multiple of \( q^* \) would work too, as \( c(q, q^*, w) \) is homogeneous of degree 0 in \( q^* \).
Each of these representations has its advantages. We will choose one or the other depending on
the context.

We also note that a behavioral demand function $c(q, w)$ cannot generally be represented by
a decision utility. Indeed, a decision utility model always generates a symmetric Slutsky matrix
$S^H(q, w)$, while it need not be symmetric in a general behavioral model. For example, in the
misperception model, with exogenous attention as in (2), $S^{H}_{ij} = S^{r}_{ij} m_j$, which is not symmetric in
general.

### 3.1.2 Planning problem

To formulate the planning problem, we introduce a social welfare function $W(v^1, ..., v^H)$ and a
marginal value of public funds $\lambda$. The planning problem is $\max_{\tau} L(\tau)$ where

$$L(\tau) = W\left((v^h(p + \tau, w))_{h=1...H}\right) + \lambda \sum_{h} \left[\tau \cdot c^h(p + \tau, w) - w\right]. \quad (23)$$

A base good, good 0, is constrained to be untaxed: $\tau_0 = 0$. Here, we assume perfectly elastic
supply with fixed producer prices $p$. We will relax this assumption in Section 5.2 where we consider
the case of imperfectly elastic supply with variable producer prices $p$.

### 3.2 Optimal Tax Formula

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

$$\gamma^h = W_{\nu^h} v^h_w + \lambda \cdot c^h_w \quad (24)$$

and $\beta^h = W_{\nu^h} v^h_w$. We also renormalize the ideal misoptimization corrective tax as:

$$\beta^h = \frac{\gamma^h}{\lambda} \quad (25)$$

We now characterize the optimal tax system.\footnote{We also have $\frac{\partial L(\tau)}{\partial \tau_i} = \sum_h (\lambda - \gamma^h) c^h_i + \beta^h D_i + \lambda \tau \cdot S^{C,s}_i$ where $S^{C}_i := S^H_i + \bar{v}_w D_i$.}

**Proposition 8 (Multiperson Ramsey, behavioral).** If commodity $i$ can be taxed, at the optimum,
$\frac{\partial L(\tau)}{\partial \tau_i} = 0$ with

$$\frac{\partial L(\tau)}{\partial \tau_i} = \sum_{h} \left[(\lambda - \gamma^h) c^h_i + \lambda(\tau - \bar{\tau}^b,h) \cdot S^{H,h}_i\right] \quad (26)$$
Proof

\[
\frac{\partial L}{\partial \tau_i} = \sum_h [W_e h v^h_{q_i} + \lambda c^h_i + \lambda \tau_i c^h_{q_i}] = \sum_h \left[ W_e h v^h_{w} q^h_{q_i} + \lambda c^h_i + \lambda \tau_i c^h_{q_i} \right]
\]

\[
= \sum_h \left[ W_e h v^h_{w} \left(-c^h_i - \tau_i \cdot S_{i}^{H,h}\right) + \lambda c^h_i + \lambda \tau_i \cdot \left[ c^h_{w} \left(-c^h_i - \tau_i \cdot S_{i}^{H,h}\right) + S_{i}^{H,h}\right] \right] \text{ using (19) and (20)}
\]

\[
= \sum_h \left[ (\lambda - \gamma^h) c^h_i + \left[ \lambda \tau_i - \gamma^h \cdot \tau_i^{h,h} \right] \cdot S_{i}^{H,h} \right] \text{ using (24)}
\]

\[
= \sum_h \left[ (\lambda - \gamma^h) c^h_i + \lambda (\tau_i - \tau_i^{h,h}) \cdot S_{i}^{H,h} \right]
\]

\[
\square
\]

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.

Let us start with the mechanical effect \(\sum_h (\lambda - \gamma^h) c^h_i d\tau_i\). If there were no changes in behavior, the government would collect additional revenues \(c^h_i d\tau_i\) from agent \(h\), which are valued by the government as \((\lambda - \gamma^h) c^h_i d\tau_i\). Indeed, taxing one dollar from agent \(h\) to the government creates a net welfare change of \(\lambda - \gamma^h\), \(\lambda\) is the value of public funds and \(\gamma^h\) 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 \(\sum_h \lambda \tau_i \cdot S_{i}^{H,h} d\tau_i\). The change in consumer prices resulting from the tax change \(d\tau_i\) induces a change in behavior \(S_{i}^{H,h} d\tau_i\) of agent \(h\) over and above the income effect accounted for in the mechanical effect. The resulting change \(\tau_i S_{i}^{H,h} d\tau_i\) in tax revenues is valued by the government as \(\lambda \tau_i S_{i}^{H,h} d\tau_i\).

Finally, let us analyze the misoptimization effect \(-\sum_h \lambda \bar{\tau}_i^{b,h} \cdot S_{i}^{H,h}\). This effect is linked to the substitution effect. If agent \(h\) were rational, the change in behavior captured by the substitution effect has 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 \(-\lambda \bar{\tau}_i^{b,h} \cdot S_{i}^{H,h} d\tau_i = -\gamma^h \cdot \tau_i^{h,h} \cdot S_{i}^{H,h} d\tau_i\) on his utility.

In the traditional model of Diamond (1975) where all agents are rational, only the mechanical and substitution effects are present, yielding \(\frac{\partial L(\tau)}{\partial \tau_i} = \sum_h [(\lambda - \gamma^h) c^h_i + \lambda \tau_i S_{i}^{H,h}]\). Adding behavioral agents introduces the following differences. First, the changes in behavior (income and substitution effects) are modified. Second, there is a new effect: the misoptimization effect.

In the misperception model (with exogenous attention) we have \(\bar{\tau}_i^{b,h} = \frac{\gamma^h}{2} \left( \tau - \tau_i^{h,h} \right)\) and formula (26) can be rewritten as:

\[
\frac{\partial L}{\partial \tau_i} = \sum_h \left[ (\lambda - \gamma^h) c^h_i + \left[ \lambda \tau_i + \gamma^h \left( \tau_i^{h,h} - \tau_i \right) \right] \cdot S_{i}^{H,h} \right].
\]  

(27)

Formula (27) fleshes out those behavioral effects which become \(\gamma^h \left( \tau_i^{h,h} - \tau_i \right) \cdot S_{i}^{H,h} d\tau_i\). They are
captured by the difference between perceived taxes \( \tau^{s,h} \) and true taxes \( \tau \), and the appearance of the behavioral Slutsky matrix \( S^H \) rather than the rational Slutsky matrix \( S^r \). It is also instructive to see the correspondence with the simple Ramsey model. The limit of small taxes corresponds to \( \lambda - \gamma^h \) small, and with one type of agent, and formula (27) becomes, to the leading order:

\[
(\lambda - \gamma^h) c_i^h + \gamma^h \tau^{s,h} \cdot S_i^{H,h} = 0,
\]

which is the generalization of (6).

**With lump-sum taxes** Suppose that the government can use a lump-sum transfer, the same for all agents. The same conditions as above apply, but there is now an additional optimality condition: \( \sum_h (\gamma^h - \lambda) = 0 \).

Suppose that agents are homogeneous, and lump-sum taxes are available, so \( \lambda = \gamma^h \). Proposition 8 implies that the optimal tax satisfies: \( \tau = \tau^{b,h} \). The optimal tax corrects the agent’s internality. If all agents are rational, \( \tau = \tau^{b,h} = 0 \).

When agents are not homogeneous, formula (25) indicates that the marginal internalities are weighted by the social marginal utility of income \( \gamma^h \). Optimal taxes put more weight on correcting the internalities of agents to which society wants to redistribute.

**Interpretation with uncertainty and several dates** Suppose that there is uncertainty and possibly heterogeneous beliefs and several dates for consumption, and complete markets. Then, our formula (26) applies without modifications, interpreting goods as a state-and-date contingent goods.\(^\text{14}\)

### 3.3 Optimal Taxation with Externalities: Pigou

We consider how taxation is modulated by externalities. Now, utility of an agent \( h \) is \( u^h (c^h, \xi) \), where \( \xi \) is a one-dimensional externality. It depends on all consumptions as \( \xi = \xi (\{c^h\}_{h=1...H}) \). For instance, we could have \( \xi = \frac{\xi}{H} \sum_h c_i^h \) if good 1 creates an externality (e.g. second-hand smoke) and \( u^h (c^h, \xi) = u^h (c^h) - \xi \). The indirect utility is now \( v^h (q, w, \xi) \).

The planning problem becomes \( \max_\tau L (\tau) \) with

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

where \( \xi = \xi (\{c^h\}_{h=1...H}) \) is endogenous to the tax system.

We define the following quantity: \( \Xi := \frac{\sum_h \left[ \beta^h \frac{u^h}{c^h} + \lambda \tau \cdot c^h \right]}{1 - \sum_h \xi \cdot c^h} \). Here \( \Xi \) is the social marginal value of the externality, including all its indirect effects on consumption as it affects tax revenues (the \( \tau \cdot c^h \)

\(^{14}\) 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.
terms) and extra “multiplier” effects on other externalities (the denominator). Hence $\Xi$ is negative for a bad externality, like pollution. We also define the ideal agent-specific Pigouvian tax $\tau^{\xi h} := -\frac{\Xi_{\cdot h}}{\lambda}$. 

It is minus the dollar value ($\Xi_{\cdot h}$) of the externality created by one more unit of consumption by agent $h$ (the $\xi_{\cdot h}$ term). We finally define $\gamma^{h,\xi}$ to be the externality-augmented social marginal utility of income:

$$\gamma^{\xi, h} := \gamma^h + \Xi_{\cdot, c_h} c_h^h = W_{w^h} v^h_{w^h} + \lambda \left( \tau - \tau^{\xi h} \right) \cdot c_w^h$$

It reflects that as one dollar is given to the agent, his direct social utility increases by $\gamma^h$, but that extra dollar changes consumption by $c_w^h$, hence the total externality by $\xi_{\cdot h} c_h^h$, so that the extra welfare impact via the externality is $\Xi_{\cdot h} c_h^h$.

We accordingly renormalize the ideal misoptimization corrective tax as:

$$\tilde{\tau}^{b, \xi, h} = \frac{\gamma^{\xi, h}}{\lambda} \cdot \tau^{b, h}.$$ 

The next proposition generalizes Proposition 8.

**Proposition 9** With an externality, the multi-person Ramsey Proposition 8 becomes:

$$\frac{\partial L(\tau)}{\partial \tau_i} = \sum_h \left[ \left( \lambda - \gamma^{h, \xi} \right) c_i^h + \lambda \left( \tau - \tau^{\xi h} - \tilde{\tau}^{b, \xi, h} \right) \cdot S^H_{i, h} \right]$$

In the case where $\xi$ depends only on total consumption ($\xi = \xi \left( \sum_h c^h \right)$), we extend the additivity result of Sandmo (1975): the externality term enters additively in the tax, as $\tau - \tau^{\xi h}$. The system of optimality conditions is the same as in the pure Ramsey case without externalities, but expect that we replace $\tau$ by $\tau - \tau^{\xi h}$ (and $\tau^{sh}$ by $\tau^{sh} - \tau^{\xi h}$). So formally at least, to the earlier optimal Ramsey taxes we must add the corrective Pigouvian taxes $\tau^{\xi h}$.

### 3.4 Optimal Nudges

In this section, we offer a model of nudges, and derive a formula for the optimal nudge in a tax system.

At an abstract level, a nudge modifies consumption as $c(p, w, \chi)$, where the nudge vector $\chi$ parametrizes the nudge. For instance, in the misperception model, a nudge changes the perceived prices according to $q^s(q, w, \chi)$.

Here is a concrete model of a nudge that anchors the consumption of a good $i$ on a default quantity $\chi$. In the absence of a nudge, the agent has decision utility $u^b$. With the nudge, he maximizes $c(q, w, m_{\chi}, \chi) = \arg\max_c u^b(c, m_{\chi}) - \Lambda B^s(q, c, m_{\chi}, \chi)$, with $B^s(q, c, m_{\chi}, \chi) = q \cdot c +$
Proposition 10: The marginal impact of a nudge $\chi$ is: \[ \frac{\partial L}{\partial \chi} = 0 \] where
\[
\frac{\partial L}{\partial \chi}(\tau, \chi) = \sum_{h} \left[ \lambda (\tau - \tau^{\text{th}}) - \frac{\beta^h}{\gamma^h, \xi^h} \tau^{h, \xi, h} \right] c^h
\]

The optimality conditions for taxes $\frac{\partial L}{\partial \tau}$ are unchanged.

This formula clarifies how best to use nudges. It has two terms corresponding to two potentially conflicting goals of nudges. The first term $\beta^h \left( \frac{u^h}{\nu^h} - q \right) c^h$ captures that nudges should be used to minimize the cost of misoptimization. The second term $\lambda (\tau - \tau^{\text{th}}) c^h$ captures that nudges should be used to increase tax revenues.

Example: a fraction $1 - f$ of the agents are rational, and can’t be nudged, but also don’t have internalities: their $\kappa_i$ is 0. A fraction $f$ of agents are nudgeable have a very high $\kappa_i$, such that $|u_{c_i}(\chi) - p| < \kappa_i$ for all such $i$; call call $N$ the set of such nudgeable agents. This means that nudged agents consume exactly $\chi$. They have $\tau^{sh} = u_{c_i}(\chi) / \lambda - p_i$. Their internality is captured by $\tau^{\text{th}} = b^h$ (if they consume too much of the good, $b^h < 0$).

Using Proposition 10, in the case with quasi-linear utility (so that $\gamma^{\text{th}} = \lambda$) we have
\[
\frac{\partial L}{\partial \chi} = \lambda \sum_{h} (\tau^{sh} - \tau^{\text{th}}) S^h_j \frac{\partial p^s}{\partial \chi} = \lambda \sum_{h \in N} \left( \frac{u_{c_i}(\chi)}{\lambda} - p - b^h \right) S^h \frac{\partial p^s}{\partial \chi}
\]
as the sum is 0 for rational agents ($\tau^{\text{th}} = 0$ as they don’t have internalities). To simplify calculations, assume that $S^h \frac{\partial p^s}{\partial \chi}$ is the same for all agent $h \in N$. The optimal nudge is the $\chi$ that satisfies: $\left( \frac{u_{c_i}(\chi)}{\lambda} - p - b^h_{h \in N} \right) = 0$. The losses from the nudge strategy are as in the case of the quantity mandate.

We remark that, as is already well understood, nudges affect only behavioral agents, so are more
refined that crude quantity regulation. On the other hand, given (here at least) nudges cannot be fine-tuned to each type of behavioral agents, they typically don’t implement the first best.

3.5 Examples

Here we extract concrete insights from Proposition 8 and its extension to externalities in Proposition 9.

3.5.1 Tax formula with constant marginal utility of wealth

This case is most useful to gain intuition. Agent’s consumption is decomposed into $c = (c_0, C)$, and utility is $u(c_0, C) = c_0 + U(C)$, while true utility is $u(c_0, C) = c_0 + U(C)$. We normalize $p_0 = q_0 = 1$. The agent chooses $C^b(q^*) = \arg \max_C U^b(C) - q^* \cdot C$, $U^b$ is “decision utility” and $q^*$ is the perceived price. Then, we define $S^r(q^*) = \frac{\partial C^{b,q}(q^*)}{\partial q^*}$ the rational Slutsky matrix of an agent with utility $U^b$ perceiving prices correctly, and $M_{ij} = \frac{\partial h^s}{\partial q_{ij}}$ the matrix of marginal perceptions, $\tau^{I,h} = U^b_C(C) - U_C(C, q) + p - p^s$ the “internality”, which could be non-zero even before taxes, and $\tau^{X,h} = \frac{\gamma^{k,h}}{\lambda} \tau^{I,h} + \tau^{k,h}$ the sum of the internality plus externality. In all those definition, we omit the row and columns corresponding to good 0, which has no distortions. We further impose $\tau^{s,h} = M^h \tau$, i.e. constant attention with zero default tax rate.

**Proposition 11** (Optimal tax formula with constant marginal utility of wealth) *In the case above, the optimal tax is:*

$$\tau = \left( \sum_h M^{h^r} S^{h,r}(I - (I - M^h) \frac{\gamma^{k,h}}{\lambda}) \right)^{-1} \sum_h [M^{h^r} S^{h,r} \tau^{X,h} - (1 - \frac{\gamma^{h,k}}{\lambda}) \tau^h] = \tau^{Pigou} + \tau^{Ramsey}$$

(29)

$$\tau^{Pigou} = \left( \sum_h M^{h^r} S^{h,r}(I - (I - M^h) \frac{\gamma^{k,h}}{\lambda}) \right)^{-1} \sum_h M^{h^r} S^{h,r} \tau^{X,h}.$$  

(30)

**Proof.** We have, from Proposition 28,

$$\tau^{b,h} = u^{b,h}_C(C^h) - u^{h}_C(C^h) + p - p^s + \tau - \tau^{s,h} = \tau^{I,h} + \tau - \tau^{s,h} = \tau^{I,h} + (I - M^h) \tau$$

hence

$$\tau - \tau^{k,h} - \frac{\gamma^{k,h}}{\lambda} \tau^h = \tau - \tau^{k,h} - \frac{\gamma^{k,h}}{\lambda} (\tau^{I,h} + (I - M^h) \tau) - \tau^{k,h} = \left( I - (I - M^h) \frac{\gamma^{k,h}}{\lambda} \right) \tau - \tau^{X,h}$$

Hence, Proposition 9 implies:

$$\sum_h \left(1 - \frac{\gamma^{h,k}}{\lambda}\right) \tau^h = - \sum_h (S^{H,h})^t \left( \tau - \tau^{k,h} - \tau^{b,k,h} \right) = - \sum_h M^{h^r} S^{h,r} \left( \left( I - (I - M^h) \frac{\gamma^{k,h}}{\lambda} \right) \tau - \tau^{X,h} \right)$$
which gives (29). □

Next, assume that $M^h$ is invertible, and define $\tau^{X,M,h} := (M^h)^{-1} \tau^{X,h}$ the ideal corrective tax for agent $h$, taking into account his misperception, and $S^{h,M,r} := M^{hr}S^{h,r}M^h$. We take the limit of a small or absent redistribution or taxation motive, $\varepsilon := \max_h \left| \frac{\gamma^{h,\xi}}{\lambda} - 1 \right|$ assumed to be small, calling also $\|\tau^{X,M}\| := \sup_h \|\tau^{X,M,h}\|$. Then, we have:

$$\tau^{Pigou} = (\sum_h S^{h,M,r})^{-1} \cdot \sum_h [S^{h,M,r} \tau^{X,M,h}] + O (\varepsilon \cdot \|\tau^{X,M}\|)$$

(31)

$$\tau^{Ramsey} = -(\sum_h S^{h,M,r})^{-1} \cdot \sum_h (1 - \frac{\gamma^{h,\xi}}{\lambda}) c^h + O (\varepsilon^2)$$

(32)

Those formula represent the optimal tax for rational agents – except that pseudo-agent $h$ has Slutsky matrix $\Sigma^{h,M,r}$ and corrective tax $\tau^{X,M,h}$. We can also write:

$$\tau^{Pigou} = E [\tau^{X,M,h}] + E [S^{h,M,r}]^{-1} \text{cov} (S^{h,M,r}, \tau^{X,M,h}) + O (\varepsilon \cdot \|\tau^{X,M}\|)$$

(33)

with $\text{cov} (S^{h,M,r}, \tau^h) = \sum_j \text{cov} (S_{ij}^{h,M,r}, \tau_j^h)$. The first term is form of “principle of targeting” (see Sandmo (1975) and Allcott, Mullainathan and Wozny (2014)). The term $E [S^{h,M,r}]^{-1} \text{cov} (S^{h,M,r}, \tau^{X,M,h})$ leads to deviations from the “principle of targeting”, as we shall now see.

### 3.5.2 Correcting int-externalities: Relaxation of the principle of targeting

We will see how misperception of taxes lead to a reconsideration of the “principle of targeting.” We assume that consumers have the same int-externality $\tau^X = \tau^{h} + \tau^l$, but heterogeneous (mis)perception of taxes, and that $\gamma = \lambda$, so there is no revenue-raison motive. Thus, (29) becomes:

$$\tau = (E [M^{hr}S^r M^h])^{-1} \cdot E [M^{hr}] S^r \tau^X.$$

When agents fully perceive taxes ($M^h = I_n$), the optimal tax is $\tau = \tau^X$: the tax exactly corrects the int-externality. With uniform misperceptions ($M^h = M$), the same applies, with an adjusted tax: $\tau = M^{-1} \tau^X$. However, when perceptions are not uniform, something else happens.

We consider the case with $n = 2$ goods, where good 1 has an int-externality, but not good 2: $\tau^X = (\xi_*, 0)$ with $\xi_* > 0$. Recall that goods 1 and 2 are substitutes (respectively complements) iff $S_{12}^r > 0$ (respectively $S_{12}^r < 0$). For instance, arguably, fuel and solar panels are substitutes, and fatty hamburgers and lean turkey are too. Propositions 12 and 13 below are reminiscent of Allcott, Mullainathan and Taubinsky (2014).

**Proposition 12** (Modified Principle of targeting) We assume that there is a negative int-externality on good 1, and none on good 2. Suppose first that agents perceive taxes correctly; then good 1 should be taxed, but good 2 should be left untaxed – that is the classic principle of targeting. Suppose
next that agents misperceive the tax on good 1 in heterogeneous ways ($\text{var}(m_{1h}) > 0$), and that $E \left[ m_{2h} - \frac{E[m_{1h}m_{2h}]}{E[m_{1h}]}m_{1h} \right] > 0$ (i.e., the perceptions of goods 1 and 2 are only mildly correlated). Then, good 2 should be subsidized (resp. taxed) if and only if goods 1 and 2 are substitutes (resp. complements).

In classical economics, if fuel pollutes, the optimal policy is to tax it, and not to subsidize substitute goods like solar panels (this example is from Salanié (2011)). Likewise, if fat is bad for consumers, and they suffer from an internality, fat should be taxed, but lean turkey should not be subsidized. Proposition 12 shows that if people have heterogeneous attention to the fuel tax, then solar panels should be subsidized. The reason is that the optimal fuel tax does not completely do its job (because of the heterogeneity of attention). So it should be completed with a subsidy to substitute goods.\footnote{A similar logic applies in the traditional model, if there is a heterogeneity in both the int-externality, and the Slutsky matrix.} Likewise, if people don’t pay attention to the fat tax, then lean turkey should be subsidized.

3.5.3 Should we tax gas or fuel inefficiency?

In the traditional model, one should just tax gas. What happens with a behavioral model, with both internality and externality? To model this, we posit that consumer $h$ needs to buy a car with fuel inefficient $c_1 = f$. As a by-product, that forces consumption of gas, $c_2 = b_h f$, where $b_h$ is an intensity of consumption. There is a tax $\tau_1$ on fuel inefficiency, and a tax $\tau_2$ on gas. Utility is $U(c_1, c_2) = -\frac{1}{2}c_1^2 + ac_1$. On net, the behavioral consumer maximizes over $c_1$ and $c_2$ (with $c_2 = b^h c_1$) the subjective value function:

$$U^{b,h}(c_1, c_2, m_2, m_3) = -\frac{1}{2}c_1^2 + a f - (p_1 + \tau_1) c_1 - (m_2^h p_2 + m_3^h \tau_2) c_2$$

The consumer only partially perceives the cost of gas ($m_2^h \leq 1$), and perceives even less the cost of the gas tax ($m_3^h \leq m_2^h \leq 1$). In addition, there is a externality, equal to $\xi c_2$.

The true value function is $v^h(c_1, c_2) = v^{b,h}(c_1, c_2, 1, 1)$, i.e. the utility function with full attention. The social value is $v^h(c_1, c_2) - \xi c_2$. The taxes here are purely redistributive and utility is quasi linear, so $\lambda = \gamma^h = 1$. We call $X^h := b^h (1 - m_2^h) p_2 + b^h \xi$, the sum of the “internality” ($b^h (1 - m_2^h) p_2$, as the consumer fails to anticipate the full cost of gas) and “externality” (the pollution $b^h \xi$), and $\psi_h := b^h m_3^h$ the demand sensitivity to the gas tax.

Which tax system should be set up? It is useful decompose $X_h = x_h + \beta_{X \psi} \psi_h$, where $\text{cov}(\psi_h, x_h) = 0$ and $\beta_{X \psi} = \frac{\text{cov}(X_h, \psi_h)}{\text{var}(\psi_h)}$ is a regression coefficient. Then, the solution is (as a few lines of calculations show, or just by direct verification)

**Proposition 13** The optimal tax system is $\tau_1 = E[ x_h ]$, $\tau_2 = \beta_{X \psi}$.\footnote{A similar logic applies in the traditional model, if there is a heterogeneity in both the int-externality, and the Slutsky matrix.}
To gain intuition, consider first the case where there only an externality, and all agents are rational. Then, \( \tau_1 = 0 \) and \( \tau_2 = \xi \) (indeed, \( X^h = b^h \xi \) and \( \psi_h = b, \) so \( \beta_{X\psi} = \xi \) and \( x_h = 0 \)). This is the classic Pigouvian case, only a tax on gas is needed as the externality comes from gas. Next, take the case with behavioral agents, \( m_3 < 1 \). When the int-externality and gas sensitivity have 0 covariance (\( \beta_{X\psi} = 0 \), perhaps because people are not sensitive to the gas tax), \( \tau_1 > 0, \tau_2 = 0 \): only fuel inefficiency of the car is taxed. When the int-externality and the sensitivity to the gas tax are perfectly correlated (for instance, because say only car usage \( b \) varies), then \( \tau_1 = 0, \tau_2 > 0 \): only the gas is targeted. Hence, gas is taxed more if it’s more correlated with the int-externality.

This example may capture non-economists’ criticism of a gas tax: it’s too “subtle” to really work. Hence, we ensure gas efficiency, when agents are naive (in a heterogeneous way), then directly targeting the “salient” good (the sticker tax on fuel inefficiency) can be a better way to proceed than a pure Pigouvian tax.

### 3.5.4 Cross-Effects of Attention

How does attention to one good affect the optimal tax on another? Let us unpack the answer from (29) in the case of a representative consumer (so that we drop the index \( h \)), and no int-externality, \( \tau^X = 0 \). Calling \( \Lambda := \frac{1-\gamma}{\gamma} \), formula (29) becomes:

\[
\tau = -\Lambda (M^r S^r M)^{-1} c
\]

which is the generalization of Proposition 1, which gave \( \tau_i = \frac{\Lambda}{m^2_i \psi_i} \).

To gain intuition, we take \( n = 2 \) goods, \( M = \text{diag}(m_1, m_2) \), normalize \( p_1 = p_2 = 1 \), and write the rational Slutsky matrix as \( S^r = -\begin{pmatrix} c_1 \psi_1 & \sqrt{c_1 c_2 \psi_1 \psi_2 \rho} \\ \sqrt{c_1 c_2 \psi_1 \psi_2 \rho} & c_2 \psi_2 \end{pmatrix} \). Calculation of \( \tau = -\Lambda (MS^r M)^{-1} c \) gives the following.

**Proposition 14** (Impact of cross-elasticities on optimal taxes with inattentive agents) When there are two goods, the optimal tax on good 1 is

\[
\tau_1 = \frac{\Lambda}{m^2_1 \psi_1} \cdot \frac{1 - \rho V_{m^1 \psi_1 c_2}}{1 - \rho^2}
\]

If goods 1 and 2 are substitutes (resp. complements), the tax on good 1 increases (resp. decreases) when \( m_2 \) falls.

To gain intuition, recall that \( \rho < 0 \) iff the goods are substitutes.\(^{16}\) When \( m_2 \) falls, the planner wants to in some sense increase the tax on good 2 by the effects in Proposition 1. To minimize distortions, he also wants to tax similar “substitute” goods, like good 1.

\(^{16}\) Also, \( \rho^2 < 1 \) as \( S^r \) is a \( 2 \times 2 \) negative definite matrix, \( 0 < \det S^r = c_1 c_2 \psi_1 \psi_2 (1 - \rho^2) \).
Perhaps curiously, we can have $\frac{\partial \tau_1}{\partial m_1} > 0$, which happens if and only if $2 < \frac{m_1}{m_2} \sqrt{\frac{\psi_{1c_2}}{\psi_{2c_1}}}$. That latter condition is quite extreme, and would imply that $\tau_1 < 0$ even though the planner wants to raise revenues. This is because the planner wants to increase consumption of the low elasticity (low $\mu_2$), good 2, he wants to subsidize good 1 if it is complement of good 2 (if $\rho > 0$).

3.5.5 Interactions between internalities and redistribution

Should we tax harmful goods, if they have a strong internality? To see the situation, take the case where a good 1 is solely consumed by a class of agents, $h$. For simplicity, assume that good 1 is separable, $u^h(c) = u^h_1(c_1) + u^h_{-1}(c_{-1})$, and the good is harmful, $\tau_{1,h} > 1$.

As a concrete example, $h$ could stand for “poor” and good 1 for “sugary sodas”. This would capture the recent policy proposal considered in New York City to tax sugary sodas.

Proposition 8 implies (with $\psi_1 := -S_{11}/c_1$): $0 = \left(\lambda - \gamma^h\right) c_1 - \lambda \left(\tau_1 - \frac{\gamma^h}{\mu_1} \tau_{h} \psi_1\right) c_1 \psi_1$, i.e. the optimal tax is:

$$\tau_1 = \frac{\lambda + \gamma^h \left(\tau_{1,h} \psi_1 - 1\right)}{\lambda \psi_1} \quad (36)$$

The sign is ambiguous, because there are two forces at work. First, because the good is harmful, it should be taxed when redistribution motives are absent ($\gamma^h = \lambda$). Second, suppose $h$ stands for “poor” and that $\gamma^h$ is high: then, if $\tau_{1,h} \psi_1 < 1$ (so that the good is not very harmful), then good 1 should be subsidized, because it’s a way to redistribute towards the poor. When the government cares more about the poor (i.e., when $\gamma^h$ is higher), the tax increases if and only if $\tau_{1,h} \psi_1 > 1$, i.e. if the internality-correction motive is stronger than the redistribution motive. This may also be why government choose to ban say high-interest rate loans, rather than tax them: taxes create an additional burden for the poor.

This situation highlights one of the advantages of nudges: nudges are a way to correct internalities without mechanically redistributing. To capture this idea, imagine the internality can be completely eliminated by an informational nudge (e.g. explaining the bad consequences of sugar), that sets $\tau_{1,h} = 0$. Then optimal taxes are used only for redistribution, $\tau_1 = \frac{\lambda - \gamma^h}{\lambda \psi_1}$.

3.5.6 Do more mistakes by the poor lead to more redistribution?

In this example, we consider the consequences for income taxation of the hypothesis that behavioral biases and incomes are negatively correlated, or in other words, that richer agents make fewer mistakes. There are two goods with prices normalized to one. The decision and experienced utility of agent $h$ are given by $u^{h,b}(c_1, c_2) = \frac{\alpha^{h,b}_1 c_1 + \alpha^{h,b}_2 c_2}{\alpha^{1}_1 \alpha^{2}_2}$ and $u^{h}(c_1, c_2) = \frac{\alpha^{1}_1 c_1^{\alpha^2} + \alpha^{2}_2 c_2^{\alpha^1}}{\alpha^{1}_1 \alpha^{2}_2}$, with $\alpha^{h,b}_1 + \alpha^{h,b}_2 = \alpha^1 + \alpha^2 = 1$.

The indirect utility of agent $h$ with (post-tax) income $z$ is then given by

$$u^h(z) = A^h z, \quad A^h = \left(\frac{\alpha^{h,b}_1}{\alpha^1}\right)^{\alpha^1} \left(\frac{\alpha^{h,b}_2}{\alpha^2}\right)^{\alpha^2} \leq 1.$$
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 \( \tau_z \) 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 \( \sum_h \frac{(\bar{z}^h)^{1-\sigma}}{1-\sigma} \), where \( \sigma \) indexes the intensity of the social preference for equality.

The planning problem is then \( \max_{\tau_z} L(\tau_z; \{A^h\}) \) with \( L(\tau_z; \{A^h\}) = \sum_h \frac{1}{1-\sigma} (A^h)^{1-\sigma} (z^h + \tau_z (\bar{z} - z^h)) \).

**Proposition 15** 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 \( \tau_z \)). 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 \( \gamma^h = (A^h z)^{-\sigma} A^h \). Indeed the marginal benefit \( \frac{\partial L}{\partial \tau_z} \) of increased redistribution (a higher tax \( \tau_z \)) is inversely related to the covariance between the social marginal utility of income and income

\[
\frac{\partial L}{\partial \tau_z} = \sum_h \gamma^h (\bar{z} - z^h) = -\text{Cov} \left( \gamma^h, z^h \right).
\]

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

### 3.6 Optimality of Simple Taxes

For simplicity, we go back to the model without Pigouvian externalities. We assume that people pay disproportionate attention to the maximum tax. To generalize the idea, we assume that they also pay attention to other quantiles of taxes, for instance the median tax. Because we think that perceptions are formed in percentage (“ad valorem”) form, we normalize here \( p_i = 1 \). Formally, we suppose that the perceived tax can be expressed:

\[
\tau_i^x (\tau, \theta) = f^i (\tau^{(r_1)} \ldots \tau^{(r_q)}, \tau, \theta)
\]  

(37)
for some function $f^i$ that is $C^1$. Here $\tau^{(r)}$ is the $r$–th quantile of taxes, for instance $\tau^{(1)}$ is the maximum tax and $\tau^{(1/2)}$ is the median tax.\footnote{Formally, $\tau^{(r)} = \inf \{ \tau \text{ s.t. } \# \{ \tau^i \geq \tau \} \geq (1-r)n \}$. This tilts the quantiles towards the higher values, for instance with that definition, the median of \{1, 2, 3, 4\} is 3.} Hence, people pay special attention to the maximum tax (the quantile 1, i.e. 100%): this is $r_1 = 1$ is one of the quantile considered. We suppose here that $f^i_{x(r_1)}$ is bounded below by a strictly positive constant.

Here $\theta$ smoothly parametrizes the attention function. We also consider a parameter $\theta_0$ such that $f \left( (\tau^{(r_1)},...\tau^{(r_Q)}), \tau, \theta_0 \right)$ does not depend on $\tau$. This means that when $\theta = \theta_0$, agents just pay attention to the quantile taxes (e.g., just to the maximum tax), but not to other taxes. When $\theta$ is close to $\theta_0$, the agent pays a lot of attention to the quantile taxes, and little to the other taxes.

One concrete example of this is:

$$\tau^a_i(\tau, \theta) = (1-m_i) \tau^d + m_i \tau_1,$$

$$\tau^d = \sum_{k=1}^{Q} \alpha_k \tau^{(r_k)} + \tau_\omega$$

with a positive weight on $\tau_{max}$, i.e. $\alpha_i > 0$. Then $M_{ij}(\tau, \theta) := f^i_j = (1-m_i) \omega_j + m_i \alpha_i$ is $\tau_i$ is different from all quantiles. Here $\theta = (m, \omega, \alpha)$, so for instance $\theta_0 = (0,0,\alpha)$ ensures $f \left( (\tau^{(r_1)},...\tau^{(r_Q)}), \tau, \theta_0 \right)$ is independent of $\tau$, and $M_{ij}(\tau, \theta_0) = 0$. So, $\theta$ close to $\theta_0$ means that $\omega$ and $m$ are small.

With $H$ types of agents, each agent $h$ has her own perception function $f^{i,h}$. We assume that the same parameter $\theta_0$ implies $f^{i,h} (\tau^{(r_1)},...,\tau^{(r_Q)}, \tau, \theta_0)$ independent of $\tau$ for all $h$.

**Proposition 16** (Simple taxes – multiperson Ramsey) We suppose that at $\theta_0$, the optimum tax features $\sum_h (\lambda - \gamma^h) c^h_i > 0$ for all $i$. Then, generically, if $\theta$ is near $\theta_0$, the optimal tax system is simple, i.e. has at most $Q$ different tax rates.

So we have simple taxes if (and essentially iff) for all $i$, $\text{cov} (\lambda - \gamma^h, c^h_i) > 0$. This condition is verified in natural cases: for instance, if the rich (high $h$) consume more of all goods than the poor (low $h$), and $\gamma^h$ decreases in $h$, then $\text{cov} (\lambda - \gamma^h, c^h_i) > 0$, for all goods.

In some cases, the condition will not hold. E.g. if the planner wants to help a lot a special segment of the population who likes opera, there will be a subsidy for opera.

### 3.7 Attention and Welfare

**Modeling attention** To capture attention and its costs, we propose the following reinterpretation of the general framework. We imagine that we have the decomposition $c = (C, m)$, where $C$ is the vector of traditional goods (champagne, leisure), and $m$ is the vector of attention (e.g. $m_i$ is attention to good $i$). We call $I^C$ (respectively $I^m$) the set of indices corresponding to traditional
The decision utility function decision heuristics is not counted in the utility function (in practice, this is largely because that in that view, people use decision heuristics, that can respond to incentives, but the cost of those additional goods that depends on attention is assume to have zero tax, we have 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 a few benchmarks. The first benchmark is “optimally allocated attention”, which we capture as follows. We suppose that there is a primitive choice function \( C(q, w, m) \) for traditional goods that depends on attention \( m = (m_1, ..., m_A) \). Attention \( m = m(q, w) \) is then chosen to maximize \( u(C(q, w, m), m) \). This generates a function \( c(q, w) = (C(q, w, m(q, w)), m(q, w)) \). The first order condition characterizing the optimal allocation of attention can be written as \( \tau^b \cdot (c_j) = 0 \) for all \( j \in I^m \). 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 \( \sum_{i \in I^c} \tau^b_i C_{m_i}(q, w, m) + \tau^b_{k(j)} = 0 \) for all \( j \in \{1, ..., A\} \).

The second benchmark is “no attention in welfare”, where attention is endogenous (given by a function \( m(q, w) \)) but its cost is assumed not to directly affect welfare so that \( u(C, m) = U(C) \). In that view, people use decisions heuristics, that can respond to incentives, but the cost of those decision heuristics is not counted in the utility function (in practice, this is largely because that cost is hard to measure). For example, in the decision vs. experienced utility model, the attention function \( m(q, w) \) could be generated by assuming that the consumer maximizes in \( C \) and \( m \) the decision utility \( u^b(C, m) \) s.t. \( \sum_{i \in I^c} q_i C_i \leq w \). The key is that attention \( m \) enters decision utility \( u^b \) but not welfare \( u \). For instance we might have \( u^b(C, m) = U^b(C) - g(m) \) for some cost function \( g(m) \). A more general function \( u^b \) might capture that attention is affected by consumption (e.g., of coffee) and attention affects consumption (by needing aspirin). Then \( \tau^b_i = 0 \) for \( i \in I^m \).

**Optimal taxation with endogenous attention** The tax formula (26) has a term \((\tau - \bar{\tau}^b_h) \cdot S^H_i \), 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 \( k \in I^m \). To think about \( \bar{\tau}^b_h \) requires no special treatment in general: this term can account for “misoptimization” terms in attention. However, two polar cases are worth considering, that simplify the calculations.

First, consider the "no attention in welfare" case. In this case we saw that \( \bar{\tau}^b_h = 0 \) for \( k \in I^m \). Together with \( \tau_k = 0 \) for \( k \in I^m \), this implies that \((\tau - \bar{\tau}^b_h) \cdot S^H_i = \sum_{k \in I^c} (\tau_k - \bar{\tau}^b_h) S^H_{ki} \) is the sum restricted to commodities.
Second, consider the “optimally allocated attention” case. Proposition 30 indicates that

\[(\tau - \tau^b) \cdot S^H_i = \sum_{k \in I^c} (\tau_k S^H_{ki} - \tau^b_k S^H_{ki|m})\]  

(38)

where \(S^H_{ki|m}\) is a Slutsky matrix holding attention constant, which is in general different from \(S^H_i\). The reason is that for welfare, it is enough to consider constant attention by the envelope theorem. For tax revenues, the full Slutsky matrix, including changes in attention matters (the term \(\tau S^H\)). However, for welfare, when attention is assumed to be optimally allocated, it is the Slutsky matrix holding attention constant that matters (the term \(\tau^b S^H\)).

4 Nonlinear Income Taxation and Mirrlees Problem

4.1 Setup

Agent’s behavior There is a continuum of agents indexed by skill \(n\), with density \(f(n)\). Agent \(n\) has a utility function \(u^n(c, z)\), where \(c\) is his one-dimensional consumption, \(z\) is his pre-tax income, with \(u_z \leq 0\). If the agent’s pre-tax wage is \(n\), and \(L\) is his labor supply, and utility is \(U^n(c, L)\), then \(u^n(c, z) = U(c, \tilde{n})\). Note that this assumes that the wage is constant (normalized to one). We discuss the impact of relaxing this assumption in Sections 5 and 14.4 in the online appendix.

If an agent earns \(z\), the government taxes her \(T(z)\), so that the disposable revenue is \(R(z) = z - T(z)\). We call \(q(z) = R'(z) = 1 - T'(z)\) the local marginal retention rate, \(Q = (q(z))_{z \geq 0}\) the ambient vector of all marginal retention rates, and \(r_0 = R(0)\) the transfer at 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.

Like in the Ramsey case of Section 3, we start with a general behavioral model to express the concepts. For simplicity, we take attention \(m\) to be exogenous and omit the dependency on \(m\) of our various functions. One could generalize our analysis to the case where attention \(m\) is endogenous, as we did in Section 3.7 for the Ramsey problem.

A misperception model However, the agent may misperceive the tax schedule, including her marginal tax rate. We call \(T^{n,s}(q, Q, r_0)(z)\) the perceived tax schedule, \(R^{n,s}(z) = z - T^{n,s}(q, Q, r_0)(z)\) the perceived retention schedule and \(q^{n,s}(q, Q, r_0)(z) = \frac{dR^{n,s}(q, Q, r_0)(z)}{dz}\) the perceived marginal retention rate. Faced with the tax schedule, the agent’s problem is

\[\mathrm{smax}_{c,z|R^{n,s}(\cdot)} u^n(c, z) \text{ s.t. } c \leq R(z)\]
Like Saez (2001), we normalize \( \xi \) to maximize the following objective function

\[
\xi \quad \text{and}
\]

The earnings \( \xi \) of agent \( n \) might be a weighted average of marginal rates, e.g. his consumption is \( c(n) = R(z(n)) \) and his utility is \( v(n) = u^h(c(n), z(n)) \).

One example of perceived tax is \( q^{n,s}(q, r_0) = q^s(q, Q, r_0) \) with \( q^s(q, Q, r_0)(z) = mq(z) + (1 - m)(\alpha q^d(Q) + (1 - \alpha) \frac{\int_0^\infty q(z)dz^s}{z}) \), where \( m \in [0, 1] \) is the attention to the true retention rate, \( \frac{\int_0^\infty q(z)dz^s}{z} \) is the average retention rate (as in “schmeduling”), and \( \alpha \in [0, 1] \). The default perceived retention rate might be a weighted average of marginal rates, e.g. \( q^d(Q) = \int q(z) \omega(z) dz \) for some weights \( \omega(z) \).

**More general model** The concrete example above motivates considering the following more general model. The primitives are the earnings function \( z^n(q, Q, r_0, r) \) which depends on the local marginal retention rate \( q \), the ambient vector of all marginal retention rates \( Q \), and the virtual income \( r \). We also obtain an indirect utility function \( v^n(q, Q, r_0, r) = u^n(qz^h(q, Q, r_0, r) + r, z^n(q, Q, r_0, r)) \).

The earnings \( z(n) \) of agent \( n \) facing retention schedule \( R(z) \) is then the solution of the fixed point \( z = z^n(q(z), Q, r(z)) \), 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

\[
\int W(v(n)) f(n) dn + \lambda \int (z(n) - c(n)) f(n) dn
\]

Like Saez (2001), we normalize \( \lambda = 1 \). We define the marginal utility of income \( g(n) = W'(v(n)) v^n(q(z(n))) \), and

\[4.2 \text{ Saez Income Tax Formula with Behavioral Agents}\]

We give a behavioral version of Saez (2001). Like him, we identify agents with their income level \( z(n) \) instead of their skill \( n \) (we write \( n(z) \)), and most of the time we leave the dependence on \( n \) and \( z \) implicit.
4.2.1 Elasticity concepts

Given an income function $z(q, Q, r_0, r)$, we define the following quantities: $\eta = q z_r(q, Q, r_0, r)$ is the income elasticity of earnings; $\zeta^u = \frac{q}{z} z_q(q, Q, r_0, r)$ the uncompensated elasticity of labor (or earnings) supply with respect to the actual marginal tax rate $q$; $\zeta^c = \zeta^u - \eta$ the compensated elasticity of labor supply with respect to the actual marginal tax $q$; $\zeta^c_{Q_z^*} = \frac{q}{z} z_{Q_z^*}(q, Q, r_0)$ is the compensated elasticity of labor supply with respect to the marginal tax $q(z^*)$, i.e. of the tax rate at a point $z^*$ different from $z$; and $\zeta^c_{s_0} = \frac{q}{z} z_{r_0}(q, Q, r_0)$. For instance, in the misperception model $q(z^*)$ affects (in general) the default tax rate, hence the perceived tax rate at earnings $z$. \(^{18}\) In the traditional model with no behavioral biases, $\zeta^c_{Q_z^*} = \zeta^c_{s_0} = 0$. All those elasticities a priori depend the agent at earnings $z$. We leave the dependence on $z$ implicit to avoid cluttering the notations too much.

Similarly to the Ramsey model, we define, evaluating at $z = z(q, Q, r_0, r)$ and $c = q z + r$, \(^{18}\)

$$\tau^b(q, Q, r_0, r) = -\frac{q u_c(c, z) + u_z(c, z)}{v_r(q, Q, r_0, r)} \tag{40}$$

which we call the ideal misoptimization corrective tax. In the traditional model with no behavioral biases, we have $\tau^b(q, Q, r_0, r) = 0$. We define the renormalized ideal misoptimization corrective tax

$$\tilde{\tau}^b(z) = g(z) \tau^b(z). \tag{41}$$

In the appendix, we prove the following modified Roy identities:

$$\frac{v_q}{v_w} = z - \frac{\tau^b_z}{q} \zeta^c, \quad \frac{v_{Q_z^*}}{v_w} = -\frac{\tau^b_z}{q} \zeta^c_{Q_z^*}. \tag{42}$$

**Particularization to the misperception model** In the misperception model, we started with a function $z(q, q^s, r)$, and the general function $z(q, Q, r_0, r)$ can be expressed concretely as $z(q(z'), Q, r) = z(q(z'), q^s(q, Q, r_0)(z'), r)$ for any earnings $z'$. As in the Ramsey case, it is useful to express behavioral elasticities as a function of an agent without behavioral biases. Call $z^r(q^s, w) = \arg \max_q u(c, z)$ s.t. $c \leq w + q^s z$ the earnings of a rational agent facing marginal tax rate $q^s$ and non-labor income $w$. Then, $z(q, q^s, w) = z^r(q^s, w')$ with $w'$ adjusts to satisfy $w' + q^s z^r(q^s, w') = w + q z^r(q^s, w')$. We call $S^r(q^s, w') = \frac{\partial z^r}{\partial q^s} (q^s, w') - \frac{\partial z^r}{\partial w'} (q^s, w') z^r(q^s, w')$ the rational compensated sensitivity of labor supply (it is just a scalar). We also define $\zeta^{cr} = \frac{q S^r}{z}$ the compensated elasticity of labor supply of the agent if he were rational.

We define $m_{zz} = q^s(q, Q, r_0)(z)$ the attention to the own tax rate, $m_{z z^*} = q^s_{Q_z^*}(q, Q, r_0)(z)$ the marginal impact on the perceived tax rate at $z$ of an increase of the tax rates at $z^*$. Then, we have

\(^{18}\) In the language of Section 3.1.1, we use income-compensation based notion of elasticity, $S^C$, rather than the Hicksian-based notion $S^H$. 

31
the following relations (see the appendix for a detailed derivation):

\[ \zeta^c = \left(1 - \eta \frac{\tau - \tau^s}{q}\right) \zeta^{cr} m_{zz}, \quad \zeta_{Q^*} = \left(1 - \eta \frac{\tau - \tau^s}{q}\right) \zeta^{cr} m_{zz^*} \]  

(43)

\[ \tau^b = \frac{\tau - \tau^s}{1 - \eta \frac{\tau - \tau^s}{q}} \]  

(44)

Hence, if the agent overestimates the tax rate \((\tau - \tau^s < 0)\), the term \(\tau^b\) is negative. Hence, loosely we can think of \(\tau^b\) as indexing an “underperception” of the marginal tax rate.

**Particularization to the decision vs. experienced utility model** In the decision vs. experienced model with decision utility \(u^b(c, z)\), then \(\zeta_{Q^*} = 0\), and \(\zeta^c\) and \(\eta\) are the elasticities associated with decision utility \(u^b\). The ideal misoptimization corrective tax is \(\tau_b = \frac{u_b u_v - u_z}{u_v}\).

### 4.2.2 Basic effects

**Impact of a change in taxes on earnings and individual utility** We first study the impact of a small change \(\delta q_{z^*}\) of the marginal retention rate at \(z^*\), and how it affects labor supply at \(z\) (e.g. via misperceptions). We simultaneously study the impact of a lump-sum (independent of \(z\)) virtual income change \(\delta K\). It will prove conceptually and notationally useful to define: \(\tilde{\zeta}_{Q^*}(z) = \zeta_{Q^*}(z) + \zeta^c(z) \delta_z (z^*)\), where \(\delta_z\) is a Dirac distribution at point \(z\). Hence, as \(\zeta_{Q^*}(z)\) was a potentially smooth function of \(z^*, \zeta_{Q^*}(z)\) is a generalized function of \(z^*\), in the sense of the theory of distributions. From now on, we mostly use our notation convention of dropping the dependency on \(z\).

**Lemma 3** (Impact of changes in taxes on behavior and welfare) *Suppose that there is a change \((\delta q_{z^*})_{z^* \geq 0}\) to marginal retention rate schedule, and a lump sum increase in revenue \(\delta K\). The impact on earnings and agent’s welfare is:*

\[ \delta z = \frac{\eta \delta K + z \int_0^{\infty} \tilde{\zeta}_{Q^*} \delta q_{z^*} d z^*}{q - \zeta^c z R''} \]  

(45)

\[ \frac{\delta v}{v_w} = \delta K - z \frac{\tau^b}{q} \left( \zeta^c R'' \delta z + \int_0^{\infty} \tilde{\zeta}_{Q^*} \delta q_{z^*} d z^* \right) \]  

(46)

In these equations, the integrals involving \(\int_0^{\infty} \tilde{\zeta}_{Q^*} \delta q_{z^*} d z^*\) should be understood in the sense of the theory of distributions as \(z \zeta^c(z) \delta q_z + \int_{z^* = 0}^{\infty} z \zeta^c_{Q^*}(z) \delta q_{z^*} d z^*\) (reintroducing in these equations...
the dependency on \( z \), leading to

\[
\delta z = \frac{\eta(z) \delta K + z \zeta^c(z) \delta q_z + \int_{z^* = 0}^{\infty} z \zeta^c_Q(z) \delta q_z * dz^*}{q(z) - \zeta^c(z) z R''(z)}
\]

\[
\frac{\delta v}{v_w} = \delta K - z \frac{\tau^b}{q} \zeta^c(z) R''(z) \delta z - z \frac{\tau^b}{q} \zeta^c(z) \delta q_z - \int_{z^* = 0}^{\infty} z \frac{\tau^b}{q} \zeta^c_Q(z) \delta q_z * dz^*.
\]

To interpret the economics of (45), start with an increase in income \( \delta K \). It has, first, an impact on labor supply: it creates a direct change in earnings supply equal to \( \frac{\tau}{q} \delta K \). The additional term \( \zeta^c z R'' \) in the denominator of (45) is more subtle, and arises from the fact that as the agent adjusts his labor supply, he experiences a different marginal tax rate (which changes as \( R'' \delta z \)), leading to an additional change in income \( \frac{\zeta^c}{q} z R'' \delta z \). The final expression solves for \( \delta z \) as a fixed point. The term \( z \int_0^\infty \overline{\zeta} Q_z(z) \delta q_z * dz^* \) reflects the impact of a change in the marginal tax rate on earnings. The difference with Saez (2001) is that it is non-zero even when the change in the tax schedule occurs at \( z^* \neq z \). This is because when agents have behavioral biases, a change of the marginal rate at \( z^* \) potentially affects the perceived tax at \( z \).

We next interpret (46). The term \( \delta K \) is a mechanical income effect, and is the only term present in the traditional model of Saez (2001). The term \( -z \frac{\tau^b}{q} \left( \zeta^c R'' \delta z + \int_0^\infty \overline{\zeta} Q_z(z) \delta q_z * dz^* \right) \) represent the welfare impact arising from changes in behavior due to the failure of the envelope theorem because of misoptimization, respectively because movements in labor supply change the marginal tax rate \( -z \frac{\tau^b}{q} \zeta^c R'' \delta z \) along the initial schedule, and because of changes in the tax schedule itself \( -z \frac{\tau^b}{q} \int_0^\infty \overline{\zeta} Q_z(z) \delta q_z * dz^* \).

**Impact of a change in taxes on social welfare** 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 we define \( H(z) = \int_z^\infty h(z') dz' \). We also define the virtual density \( h^*(z) = \frac{1-T''(z)}{1-T''(z)+\zeta^c z R''(z)} h(z) \) which can also be written as \( \frac{1-T''(z)}{1-T''(z)+\zeta^c z R''(z)} h(z) \).

**Lemma 4** Under the conditions of the Lemma 3, the change in social welfare associated with the agent, \( \delta L(z) = -\delta K + T''(z) \delta z + g(z \frac{\delta v}{v_w}) \) is:

\[
\delta L(z) = (\gamma(z) - 1) \delta K + \frac{T''(z) - \overline{T'}(z) h^*(z)}{1-T''(z)} h(z) z \int_0^\infty \overline{\zeta} Q_z(z) \delta q_z * dz^*
\]

where \( \gamma(z) \) is the social marginal utility of income:

\[
\gamma(z) = g(z) + \eta(z) \frac{\overline{T'}(z)}{1-T''(z)} + \eta(z) \frac{T''(z) - \overline{T'}(z) h^*(z)}{1-T''(z)} h(z)
\]

This definition of the social marginal utility of income \( \gamma(z) \) is similar to the one we encountered in the Ramsey problem. It encompasses the direct impact of one extra dollar on the agent's
welfare (the \( g(z) \) term) and the impact coming from a change in labor supply on tax revenues
\( \frac{T'(z)}{1-T'(z)} \eta(z) \frac{h^*(z)}{h(z)} \). Compared to Saez (2001), it features a new term arising from the failure of the
envelope theorem, \( \eta(z) \frac{T'(z)}{1-T'(z)} \left( 1 - \frac{h^*(z)}{h(z)} \right) \).

The effect on welfare \((47)\) is much like in the multiperson Ramsey of Proposition 8. The term
\(- (1 - \gamma(z)) \delta K\) is a mechanical effect, abstracting from changes in behavior. As the government
gives back \( \delta K \) to agent, the impact on revenues is \(- \delta K\), while the impact on the agent is valued as
\( \gamma(z) \delta K\). Next, there is a substitution effect \( \frac{T'(z)}{1-T'(z)} \frac{h^*(z)}{h(z)} \int_0^\infty \xi_{Q^*} (z) \delta q^* dz^*\): as the agent changes his
labor supply, there is a change in tax revenues proportional to \( \frac{T'(z)}{1-T'(z)} \int_0^\infty \xi_{Q^*} (z) \delta q^* dz^*\). Third,
there is a misoptimization term, \( \frac{-\gamma(z) h^*(z)}{1-T'(z) h(z)} \int_0^\infty \xi_{Q^*} (z) \delta q^* dz^*\).

We also note the following first order condition for the intercept of the tax schedule.

**Lemma 5** At the optimum,

\[
\int_0^\infty \left( 1 - \gamma(z) - \frac{T'(z) - \overline{r}^b(z) h^*(z)}{1 - T'(z) h(z)} z \zeta_{Q^*} (z) \right) h(z) \, dz = 0
\]  

**Proof** Using Lemma 4, applied to a change \( \delta r_0 \) to all agents, and slightly generalizing, we find

\[
\delta L (z) = (\gamma(z) - 1) \delta r_0 + \frac{T'(z) - \overline{r}^b(z) h^*(z)}{1 - T'(z) h(z)} z \zeta_{Q^*} (z) \delta r_0
\]

and \( \delta L = \int \delta L (z) h(z) \, dz \) should be 0. \( \square \)

**Proposition 17** (Impact of a local change on the marginal tax rate on the government’s objective function) We have, with \( \frac{\partial L}{\partial \tau_{z^*}} \equiv - \frac{\partial L}{\partial q_{z^*}} \)

\[
\frac{\partial L}{\partial \tau_{z^*}} = \int_{z^*}^\infty \left( 1 - \gamma(z) \right) h(z) \, dz - \zeta_{Q^*}(z^*) \frac{T'(z^*) - \overline{r}^b(z^*)}{1 - T'(z^*)} z^* h^*(z^*) - \int_0^\infty \zeta_{Q^*}(z) \frac{T'(z) - \overline{r}^b(z)}{1 - T'(z)} z h^*(z) \, dz
\]  

This equation involves an equality between two generalized functions of \( z^* \). This is the income
tax equivalent of the formula in Proposition 8 for the multiperson Ramsey. The three
terms in \((50)\) correspond to the by now familiar mechanical \( \int_{z^*}^\infty (1 - \gamma(z)) h(z) \, dz \), substitution
\(- \zeta_{Q^*}(z^*) \frac{T'(z^*) - \overline{r}^b(z^*)}{1 - T'(z^*)} z^* h^*(z^*)\) and misoptimization \( \zeta_{Q^*}(z^*) \frac{T'(z)}{1-T'(z)} z h^*(z) \, dz \) effects. The first two terms are exactly as in Saez (2001), the third one is new as it is present only
with behavioral agents. We will describe its meaning shortly. We also note that formula \((50)\) can be written in a more compact way as:

\[
\frac{\partial L}{\partial \tau_{z^*}} = \int_{z^*}^\infty (1 - \gamma(z)) h(z) \, dz - \int_0^\infty \frac{\overline{r}^b(z)}{1 - T'(z)} z h^*(z) \, dz.
\]  

34
4.2.3 Optimal income tax formula

**Proposition 18** The optimal tax rate satisfies the formula: for all $z^*$,

$$
\frac{T'(z^*) - \tilde{\tau}^b(z^*)}{1 - T'(z^*)} = \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^\infty (1 - \gamma(z)) \frac{h(z)}{1 - H(z^*)} dz \quad (52)
$$

$$
- \int_0^\infty \frac{\zeta^c(z^*)}{\zeta^c(z)} \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} \frac{zh^*(z)}{z^* h^*(z^*)} dz. \quad (53)
$$

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

$$
\frac{T'(z^*) - \tilde{\tau}^b(z^*)}{1 - T'(z^*)} + \int_0^\infty \omega(z^*, z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} dz = \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^\infty e^{-\int_{z'}^{z^*} \lambda d s} \left( 1 - g(z) - \eta \frac{\tilde{\tau}^b(z)}{1 - T'(z)} \right) \frac{h(z)}{1 - H(z^*)} dz \quad (54)
$$

$$
= \frac{\zeta^c(z^*)}{\zeta^c(z)} \left( \int_{z^*}^{\infty} e^{-\int_{z'}^{z^*} \lambda d s} \lambda(z') \frac{\zeta^c(z')}{\zeta^c(z^*)} dz' \right) \frac{zh^*(z)}{z^* h^*(z^*)}. \quad (55)
$$

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

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

The intuition for formula (54) is as follows. The first term on the right-hand side 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 (48)). The second term 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. First, increasing the marginal tax rate at $z^*$ leads the agents at another income $z$ to perceive higher taxes on average (if $\zeta^c_{Q^*}(z) > 0$), which leads them to decrease their labor supply and reduces tax revenues (as captured by $\zeta^c_{Q^*}(z) \frac{h(z)}{h^*(z^*)}$). Ceteris paribus, this consideration pushes towards a lower tax rate, compared to the Saez optimal tax formula. Second, suppose for concreteness that the agent at $z$ overestimates the tax (so $\tilde{\tau}^b(z) < 0$). Increasing the marginal tax rate at $z^*$ leads the agent to overestimate the tax even more, which worsens the agent’s welfare. This again pushes towards a lower tax rate.

The modified Saez formula (54) uses the concept $g(z)$ rather than $\gamma(z)$ and is easily obtained from formula (52) using equation (48)). When there are no income effects ($\eta = \lambda(z) = 0$), the optimal tax formula (52) and the modified Saez formula (54) coincide. It also coincides with the traditional Saez formula (Saez 2001) when there are no behavioral biases ($\zeta^c_{Q^*}(z) = \omega(z^*, z) = \tilde{\tau}^b(z) = 0$). In this case, the left-hand side of (54) is simply $\frac{T'(z^*)}{1 - T'(z^*)}$ so that the formula solves for the optimal marginal tax rate $T'(z^*)$ at $z^*$. Of course, the formula is expressed in terms of endogenous objects or “sufficient statistics”: welfare weights $g$, 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 ideal misoptimization corrective tax \( \hat{\tau}^b(z) \) and the cross elasticities \( \zeta^{Q_z}_{\alpha} (z) \). Second, it is not possible to solve-out the optimal marginal tax rate in closed form. Instead, the modified Saez formula (54) at different values of \( z^* \) form a system of linear equations in the optimal marginal tax rates \( T'(z) \) for all \( z \).

In the case where behavioral biases can be represented by a decision vs. experienced utility model, the modified Saez formula takes a particular simple form since in this case \( \omega (z^*, z) = 0 \) and \( \hat{\tau}^b (z) = g (z) \frac{u_w}{u^w_{\tau}} - u_z \), and there is no linear system to invert.

### 4.2.4 Top marginal tax rate

We revisit the Saez analysis and study the impact of misperceptions on the top marginal tax rate. We consider the misperception version of the model, and take \( \tau^s(z) = m \tau(z) + (1 - m) \tau^d \), with \( \tau^d = \int_{z > 1} w(z) \tau(z) dz \), where \( \int w(z) dz = 1 \). We write \( w(z) = zh(z) G(z) \) for a function \( G(z) \geq 0 \). Then, \( m_{zz} = m, m_{zz^*} = \frac{d\tau^*(z)}{dz} = (1 - m) w(z^*) \). We further particularize to the case \( n = 0 \), so that \( \zeta^c (z) = \zeta^{cr} (z) m, \zeta^c_{Q_z} (z) = \zeta^{cr} (z) (1 - m) w(z^*) \) and \( \tau^b = (1 - m) (\tau - \tau^d) \).

Then, the Saez formula implies the following. We call \( \pi \) the Pareto exponent of the distribution \( h^* \), and suppose that \( \zeta^{cr} (z) \) is constant across agents. We suppose that \( G(\infty) := \lim_{z \to \infty} G(z) \) and \( \zeta^c (\infty) = \lim_{z \to \infty} \zeta^c (z) \) are finite and strictly positive.

**Example 1** (Top marginal tax rate) With a Rawlsian welfare function, the asymptotic marginal tax rate for top incomes, \( \tau^r \), satisfies:

\[
\frac{\tau}{1 - \tau} = \frac{1}{\zeta^c (\infty) \pi} - \frac{1 - m}{m} \int_{z = \infty}^{\infty} \zeta^c (z) \frac{G(\infty) h^z (z)}{G(z) h(z)} \frac{T'(z)}{1 - T'(z)} w(z) dz
\]

When \( m < 1 \) is close enough to 1, the second term on the right hand side is positive, so that the asymptotic marginal tax rate is less than its value \( \frac{1}{\zeta^c (\infty) \pi} \) in the traditional Saez analysis.

The economics is very simple. The top marginal tax rate affects not only top earners, but also (via the default tax \( \tau^d \)) the tax perceived by agents at all points of the income distribution (hence the corresponding tax revenues, after multiplication by \( T'(z) \)). The top tax rate is more salient (i.e. when \( G(\infty)/G(z) \) is higher), the optimal top rate is lower. Concretely, if France increases the top rate to 75\%, this contaminate the perception earners even below the top incomes.

As a simple back-of-envelope calculation, imagine that \( \frac{\zeta^c (z) G(\infty) h^z (z)}{\zeta^c (\infty) G(z) h(z)} = 1 \),

\[
\frac{\tau}{1 - \tau} = \frac{1}{\zeta^c (\infty) \pi} - \frac{1 - m}{m} E^w \left[ G(\infty) h^z (z) \right] E^w \left[ \frac{T'(z)}{1 - T'(z)} \right]
\]

where \( E^w [\cdot] \) is a weighted average with weights \( w(z) \) and \( w^* (z) = \frac{w(z) G(\infty)}{E^w \left[ \frac{G(z)}{G(z)} \right]} \).
Take the typical Saez calibration with $\zeta^c(\infty) = 0.2$, and $\pi = 2$. If the typical tax is $T'(z) \simeq \frac{1}{3}$ so that $E^w_w \left[ \frac{T(z)}{1-T(z)} \right] \simeq \frac{1}{2}$. With rational agents, the traditional rate is $\tau = 71\%$. When agents are behavioral with attention $m = 0.8$, the top rate drops to $\tau = 70\%$ when the typical relative salience of top marginal tax rates is $E^w_w \left[ \frac{\zeta^c(\infty)}{G(z)} \right] = 1$, and to $\tau = 56\%$ when $E^w_w \left[ \frac{\zeta(\infty)}{G(z)} \right] = 10$. This illustrates the key importance of the relative salience of the top tax rate, a sufficient statistic that plays no role in the traditional analysis.

The behavioral model could in principle allow for situations where the optimal top tax rate would be higher than under the Saez formula. For instance, that would be the case if we had $G(\infty) < 0$, i.e. as a “contrast effect”, in which people feel less the pain of taxes if the rich pay more.

We also revisit the classic result that if the income distribution is bounded at $z_{\text{max}}$, 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. The modified Saez formula then prescribes $T'(z_{\text{max}}) = \bar{\tau}^b(z_{\text{max}})$ 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).

### 4.2.5 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. This is an immediate consequence of the Saez formula. 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 $u^b$ and assume that $u^b$ is quasilinear so that there are no income effects $u^b(c, z) = c - \phi(z)$. We take experienced utility to be $u(c, z) = \theta c - \phi(z)$ with $\theta > 1$. Then the modified Saez formula becomes

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

where yielding $\bar{\tau}^b(z) = -g(z) \phi'(z) \frac{2-\theta}{\theta}$. When $\theta > 1$, we have $\bar{\tau}^b(z^*) < 0$ 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 as $\theta > 1$. Such biases could be particularly relevant at the bottom of the income distribution (see Chetty and Saez (2013) for a review of evidence). If these biases are strong enough, then 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. In independent work, Lockwood and Taubinsky (2015) derive a modified Saez formula in the context of decision vs. experienced utility model. They zoom in on the EITC insight and provide an empirical analysis and calibration.
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 appendix for a behavioral treatment of the Saez (2002) extensive margin of labor supply model.

4.3 Optimality of Simple Income Taxes

We use the notations and concepts from the treatment of simple taxes for the multi-person Ramsey case, section 3.6. We suppose that people pay disproportionate attention to the maximum tax, and possibly other quantiles of taxes. Formally, we suppose that the perceived tax can be expressed:

\[ q^*(q, Q, r_0, \theta) = f(q^{(r_1)} ..., q^{(r_Q)}, q, Q, r_0, \theta) \]

for some function \( f \) that is \( C^1 \). We assume that people do pay attention to the maximum tax (the quantile 1, i.e. 100%): this is \( r_1 = 1 \) is one of the quantile considered. We suppose here that \( f(q^{(r_1)} ..., q^{(r_Q)}, q, Q, r_0, \theta) \) does not depend on \( (q, Q, r_0) \). One concrete example of this is:

\[ q^*(q, Q, r_0, \theta) = (1 - m) q^d + m q \]

with \( q^d = \sum_{k=1}^{Q} \alpha_k q^{(r_k)} + \gamma \), with \( \alpha_1 > 0 \). Then, \( \theta = (m, \omega, \alpha) \) and \( \theta_0 = (0, 0, \alpha) \).

Suppose that \( \theta = \theta_0 \), and define \( \Gamma(z^*) = \int_{z^*}^{\infty} (1 - \gamma(z)) h(z) \, dz \). Lemma 5 ensures \( \Gamma(0) = 0 \).

We suppose that \( \Gamma(z^*) > 0 \) for all \( z^* > 0 \). This condition is natural. For instance, when \( \eta = 0 \), \( \gamma(z) = g(z) \). As the social welfare weight \( g(z) \) is decreasing in \( z \), and \( \Gamma(0) = 0 \), we will have \( \Gamma(z^*) > 0 \) for all \( z^* > 0 \).

**Proposition 19** (Simple taxes – income tax). We suppose that at \( \theta_0 \), the optimum tax features: \( \Gamma(z^*) > 0 \) for all \( z^* > 0 \). Then, generically, if \( \theta \) is near \( \theta_0 \), the optimal tax system is simple, i.e. has at most \( Q \) different marginal rates.

5 Extensions

5.1 Salience as a policy choice

Governments have a variety of ways of making a particular tax more or less salient. For example, Chetty et al. (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 how the
optimal salience of the tax system.

We investigate this question in the context of two simple examples. We start with the basic Ramsey model developed in Section 2 with $\tau^d = 0$. 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 $m$ (respectively $m'$). 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 distortive because the associated elasticities are lower $m_i \psi_i < m'_i \psi_i$.

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 5. 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 increases average attention for all agents but also decreases the heterogeneity in perceptions. We formalize this comparative static with the following two requirements $E[m'^h_i] > E[m^h_i]$ and $\frac{\text{Var}[m'_h]}{E[m'^h_i]} < \frac{\text{Var}[m^h_i]}{E[m^h_i]}$. The gain from the decreased salience is given by $L - L' = \Delta L$ with

$$\Delta L = -\frac{1}{2} \Psi H \xi^2 \left( \frac{\text{Var}[m^h_i]}{E[m^h_i]} - \frac{\text{Var}[m'^h_i]}{E[m'^h_i]} \right) < 0.$$ 

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 $S_i^{H,h}$ 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 $c(w, p, \tau_1, \tau_2, ..., \tau^K)$, where $\tau^K$ are tax vectors with different degrees of salience. To each tax instrument $\kappa$ corresponds a Slutsky matrix $S_{ij}^{H,\kappa}$ which depends on the tax instrument indexed by $\kappa$. The optimal tax formula can then be written as $\frac{\partial L(\tau)}{\partial \tau_i^\kappa} = 0$ where

$$\frac{\partial L(\tau)}{\partial \tau_i^\kappa} = \sum_h \left[ (\lambda - \gamma^h) e_i^h + \lambda (\bar{\tau} - \tau_i^h) \cdot S_i^{H,\kappa,h} \right],$$

39
with \( \bar{\sigma} = \sum_{k=0}^{K} \sigma^k \). The intuition for this formula is that the different tax instruments lead to different substitution effects captured by different Slutsky matrices \( S_{ij}^{H,\kappa} \). Note that the substitution effect associated with one tax instrument \( \kappa \) affects the common tax base of the other tax instruments \( \kappa' \), which explains why the formula features the total \( \bar{\sigma} \) rather than the individual tax \( \sigma^\kappa \). 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.\(^{19}\)

### 5.2 Supply Elasticities and Production (In)efficiency

#### 5.2.1 Optimal taxes with given efficient production

We So far we have assumed a perfectly elastic production function (constant production prices \( p \)). 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 \( \tau \).

In behavioral models, this result must be qualified. This section therefore generalizes the model to imperfectly elastic production function (non-constant prices \( p \)).

In behavioral models, prices \( p \) and taxes \( \tau \) might affect behavior differently. We introduce a distinction between taxes \( \tau^p \) that affect behavior like prices and taxes \( \tau^c \) that affect behavior differently from prices. For example \( \tau^p \) could represent taxes that included in listed prices \( p + \tau^p \) (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 \( \tau^c \) that are not included in listed prices. An agent’s demand function can then be written as \( c^h(p + \tau^p, \tau^c, w) \). This distinction will prove to be important for the generalization of our results to imperfectly elastic production functions.

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

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

\(^{19}\)Consider for example the case where there is only one agent and only one (taxed) good. With \( m' > m \), we get

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

where \( \bar{\sigma} \) is the total perceived tax arising from the joint perception of the two tax instruments. This requires \( \lambda = \gamma \) and with \( \bar{\tau} = 0 \). In other words, the solution is the first best. This is because a planner can replicate a lump sum tax by combining a tax \( \tau \) with low salience \( m \) and a tax \( -\tau \frac{m'}{m} \) with high salience \( m' > m \), generating tax revenues \( \tau \frac{m'}{m} \) 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 misperceptions of the two taxes, the first best might is not achievable.
We denote by $\bar{\tau} = \tau^c + \tau^p$ the sum of the tax vectors.

We can write the planning problem as

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

subject to the resource constraint

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

and the competitive pricing condition

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

The competitive pricing equation is a fixed point in $\mathbf{p}$. We denote the solution by $\overline{\mathbf{p}}$.

We denote the solution by $\mathbf{p}$. The derivatives of this function encapsulate the incidence of taxes depending on the demand and supply elasticities. We denote by $\varepsilon_{ji}^\kappa = \frac{\partial p_j}{\partial \tau_i^\kappa}$ the derivative of the prices $p_j$ of commodity $j$ with respect to the tax $\tau_i^\kappa$ with $\kappa \in \{p, c\}$. We note that the $\varepsilon_{ji}^\kappa$ solve the system of equations

$$F''(y)^{-1} \cdot \varepsilon_i^\kappa - \sum_{h,j} c^h_{ji} \varepsilon_{ji}^\kappa = \sum_h c^h_{i},$$

so that the $\varepsilon_{ji}^\kappa$ reflect both demand elasticities (the terms $\sum_h c^h_{ji}$) and supply elasticities (the terms in $F''(y)^{-1}$). This formula is the behavioral extension of the standard incidence calculations determining how the burden of taxes is shared between consumers and producers.

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

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

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

**Proposition 20** With an imperfectly elastic production function, the following results hold. If there is a full set of commodity taxes $\tau^p$ then the optimal tax formulas can be written as

$$0 = \sum_h \left[ (\lambda - \gamma^h) c^h_i + \lambda (\bar{\tau} - \tau^h_i) \cdot \mathcal{E}_{i}^{H,\kappa,h} \right],$$

are independent of production elasticities, and coincide with those of Section 3 if taxes are restricted to be of the $\tau^p$ type or with those of Section 5.1 if taxes can be both of the $\tau^p$ type and the $\tau^c$ type.
Second, when there is a restricted set of commodity taxes $\tau^p$, then the optimal tax formulas can be written as

$$0 = \sum_h [(\lambda - \gamma^h) c_i^h + \lambda (\bar{\tau} - \bar{\tau}^{h,h}) \cdot S_i^{H,k,h}] + \sum_h \sum_j [(\lambda - \gamma^j) c_j^h + \lambda (\bar{\tau} - \bar{\tau}^{h,h}) \cdot S_j^{H,k,h}] \varepsilon_{ji},$$

depend on production elasticities, and do not coincide with those of Sections 3 and 5.1.

With a full set of commodity taxes $\tau^p$, we can rewrite the objective function and the resource constraint in the planning problem as a function of $\theta = \tau^p + \pi$. 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 $\tau^p$ given the desired value of $\theta$. As a result, only the first derivatives of the production function $\pi = F'$ 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 $\tau^p$, 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 $\tau^p$ form. The difference is that even with a full set of commodity taxes of the $\tau^p$ (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 $\tau^p$ 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. This is developed in Proposition 47 of the online appendix.

To illustrate Proposition 20, consider the separable case $u(c) = c_0 + u(c_1)$ in the misperception case with $\tau_1^s = \tau_1^p + m_1 \tau_1^c$, with $0 \leq m_1 \leq 1$ and $\tau_1^c$ is exogenous (perhaps set to 0).

We represent the production function as follows: it takes $C(y_1)$ units of good 0 to produce $y_1$ units of good 1. We define supply and demand to be $S(p_1) = C'^{-1}(p_1)$ and $D(p_1 + \tau_1^p + m_1 \tau_1^c) = u'^{-1}(p_1 + \tau_1^p + m_1 \tau_1^c)$. We denote by $\varepsilon_S > 0$ and $\varepsilon_D > 0$ the corresponding supply and demand elasticities (corresponding to a fully perceived change in $p_1$). 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 $\tau^c$ satisfies
\[
0 = \left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} m_1 \right) + \left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} \right) \varepsilon^{c}_{11},
\]
which we can rewrite as
\[
\frac{\Lambda}{\psi_1} = \frac{\tau^p + m_1 \tau^c \varepsilon^{c}_{11}}{p_1} \frac{m_1 + \varepsilon^{c}_{11}}{1 + \varepsilon^{c}_{11}}.
\]
As long as $m_1 < 1$, the higher is the supply elasticity $\varepsilon_S$, the more the burden of the tax is shifted to the consumer, the higher is $\varepsilon^{c}_{11} < 0$, the lower is the optimal tax.

5.2.2 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 $\tau^c$ type (which would be enough to guarantee production efficiency in the traditional model), as long as there is not a full set of commodity taxes of the $\tau^p$ type.

We start by considering the case where there is a full set of commodity taxes of the $\tau^p$ type and show that production efficiency holds under some extra conditions. We denote by $\theta = \tau^p + \frac{\psi^c}{\psi^p} \tau^p \varepsilon^{c}_{11}$.

We follow Diamond and Mirrlees (1971) and establish that production efficiency holds by assuming that the planner can control production, and showing that the planner chooses an optimum on the frontier of the production possibility set. The corresponding planning problem is
\[
\max_{q, \tau^c} W \left( \left( v^h (q, \tau^c, w) \right)_{h=1\ldots H} \right)
\]
subject to the resource constraint
\[
F \left( \mathbf{g} + \sum_h e^h (q, \tau^c, w) \right) \leq 0.
\]

Proposition 21 With a full set of commodity taxes $\tau^p$, then production efficiency holds if either: (i) there are lump sum taxes and for all $q$, $\tau^c$ and $w$, $v^h_w (q, \tau^c, w) \geq 0$ for all $h$ with a strict inequality for some $h$; or (ii) for all $q$, $\tau^c$ and $w$, there exists a good $i$ with $v^h_i (q, \tau^c, w) \leq 0$ for all $h$ with a strict inequality for some $h$.

The proof is almost identical to the original proof of Diamond and Mirrlees (1971). Note however

---

20 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 $\left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} m_1 \right) = 0$. Now imagine that $\varepsilon_S < \infty$. Then at this tax $\left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} \right) < 0$ so that $\left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} \right) \varepsilon^{c}_{11} > 0$ and by implication $\left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} \right) + \left( \Lambda_{c_1} - \tau^c \psi_1 \frac{c_1}{p_1} \right) \varepsilon^{c}_{11} > 0$ This implies that increasing the tax improves welfare.
that the conditions $v^{b}_{w}(q, \tau^{c}, w) > 0$ or $v^{b}_{q}(q, \tau^{c}, w) < 0$ 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 $v^{b}_{w}(q, \tau^{c}, w) < 0$. Loosely speaking, this happens if mistakes get worse as income increases. Similarly, it is entirely possible that $v^{b}_{q}(q, \tau^{c}, w) > 0$ for all $\tau$ since Roy’s identity does not hold ($\frac{\partial u}{\partial \tau} \neq -c_i$). Failures of production efficiency could then arise even with a full set of commodity taxes $\tau^p$. 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 $\tau^p$, even if there is a full set of commodity taxes $\tau^c$. Consider the following example. There are two consumptions goods, $0$ and $1$, two types of labor, $a$ and $b$, a representative agent with decision utility $u^{b}(c_0, c_1, l_a, l_b) = c_0 + U^{b}(c_1) - l_a - l_b$, and experienced utility $u^{b}(c_0, c_1, l_a, l_b) = u(c_0, c_1, l_a, l_b) - \xi_1 c_1$, where $\xi_1 > 0$ 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: $y_i = \left( \frac{l_{a_i}}{\alpha_i} \right)^{\alpha_i} \left( \frac{l_{b_i}}{1-\alpha_i} \right)^{1-\alpha_i}$, with $\alpha_i \in (0, 1)$. As before, $0$ is the untaxed good, $\tau_0 = 0$. The government can set taxes $\tau_1$, $\tau_a$, and $\tau_b$ 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 $\tau_a$, $\tau_b$, 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 $\tau_{1a}$ for the use of input $a$ in the production of good $1$. Note that production efficiency is equivalent to $\tau_{1a} = 0$.

**Proposition 22** If the consumer is fully inattentive to the tax $\tau_1$, then the optimal tax system features production inefficiency: $\tau_{1a} > 0$. If the consumer is fully attentive to the tax $\tau_1$, 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 $\tau_1$ on good $1$, hence 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.

6 Discussion

6.1 Novel Sufficient Statistics

Operationalizing our optimal tax formulas (26) and (54) 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 $\lambda$, the social marginal utilities of income
γ^h, consumption vectors \( \zeta^h \), Slutsky matrices \( S^{H,h} \), and marginal internalities as measured by the ideal misoptimization corrective taxes \( \bar{\tau}^{b,h} \). All these sufficient statistics are present in the optimal tax formula of the traditional model with no behavioral biases, with the exception of marginal internalities \( \bar{\tau}^{b,h} \). They are routinely measured by empiricists.

The marginal internalities \( \bar{\tau}^{b,h} \), 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 \( \tau^e,h \) to calibrate corrective Pigouvian taxes in the traditional 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 internalities (see Bernheim and Rangel 2009 and Mullainathan, Schwartzstein and Congdon 2012).

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 a representation by a utility function \( u \) with corresponding indirect utility function \( \nu \). Differences in choices in environments where behavioral biases are present would then allow to measure the marginal internalities \( \tau^{b,h} = q - \frac{\nu^b}{\nu^w} \). For example, if the biases arise from the misperception of taxes so that \( \tau^{b,h} = \tau - \tau^{s,h} \), then perceived taxes \( \tau^{s,h} \) 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 \( u^h \). To the extent that agents are sophisticated and understand that they are subject to these biases, experienced utility could be recovered by confronting agents with the possibility of restricting their later choice sets.

Another strategy (see e.g. Chetty, Kroft and Looney (2009) and Allcott et al. (2012)) to establish and isolate behavioral biases is 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 misperception is to use surveys to directly elicit perceived taxes \( \tau^{s,h} \) (see e.g. Liebman and Zeckhauser (2004) and Slemrod (2006)).

Similarly, in the Mirrlees model, our optimal tax formula (54) features the distribution of income \( h(z) \) and \( h^*(z) \), the social marginal utilities of income \( \gamma(z) \), the elasticities \( \zeta^c(z) \) and \( \zeta^c_{Q^*}(z) \) as well as \( \bar{\tau}^b(z^*) \). All these sufficient statistics are present in the optimal tax formula of the traditional model with no behavioral biases, with the exception of \( \zeta^c_{Q^*}(z) \) and \( \bar{\tau}^b(z) \). Even though measuring \( \zeta^c_{Q^*}(z) \) is typically neglected (because \( \zeta^c_{Q^*}(z) = 0 \) in the traditional model with no behavioral biases), measuring the elasticity \( \zeta^c_{Q^*}(z) \) poses no conceptual difficulty as it can be recovered from observation of private choices. As for \( \bar{\tau}^b(z) \), the same strategies that were discussed for \( \tau^{b,h} \) in

---

21 The Slutsky matrix is a behavioral one. In particular, it reflects misperceptions about taxes, as in (22).

22 That might help bridge the gap between micro and macro elasticities, as people are both influenced by their “local-micro” tax \( q(z) \) and the “ambient” tax rates \( q^d \), perhaps the average tax rate.
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 expectations (in our model, this is the matrix \( M_{ij} = \frac{\partial q_i(q,w)}{\partial q_j} \)).

It would be interesting to study if the condition for the optimality of simple taxes is verified in practice, and see if the maximum tax has an important influence on behavior. We argued is it. 23 The same idea holds for marginal tax rates. Is there a special impact of the top tax? In the optimal tax formula, \( \zeta_{Q_*} := \frac{\partial \ln z}{\partial Q_*} \) is important. For instance, this is the influence of labor supply by people earning $100k when the tax rate on those at $1M increases. This kind of consideration is present in the policy debate (under headings like “business-friendliness” of the environment).

6.2 Are our effects robust to learning?

Arguably they are. First, there seems to be a persistence confusion. Slemrod (2006) argues that Americans overestimate on average the odds their inheritance will be taxed. People seem to perceive average for marginal tax rates (Liebman and Zeckhauser 2004). People may overestimate the odds they’ll move to a higher tax bracket (Benabou and Ok 2001).

Second, our framework works for decision utility. There, learning may be quite slow. For instance, people may persistently smoke too much, perhaps because of hyperbolic discounting (Laibson 1997).

Third, there is a more analytical answer, developed in the Section 14.2 of the online appendix. Suppose that the government sets taxes \( \tau \) at time 0. People first are inattentive for an amount of time \( 1 - M \), and attentive for a time \( M \) (this could be the “discounted” amount of time, e.g. if it takes \( T \) years to learn, \( M = \int_0^\infty e^{-rt}dt = e^{-rT} \). Then, the loss function is:

\[
L(\tau) = -\frac{1}{2} \sum_i \left[(1 - M)(\tau_i^*)^2 + M\tau_i^*\right] \psi_i y_i + \lambda \sum_i \tau_i y_i .
\]

Hence, this is exactly like a one-period model with a fraction \( M \) of fully attentive agents. We use \( \tau_i^* = \tau_{\text{max}} \) to make the point.

**Proposition 23** (Simple tax scheme from bounded rationality — with learning) *If learning takes too long, then the optimal tax system is to have simple taxes (as in Proposition 58) forever.*

The online appendix also develops a model where the government can continuously update the tax system, and the conclusion is qualitatively the same. The idea is that complex taxes have short-run costs (as people make mistakes) and long-run benefits (as the system is closer to the Ramsey ideal when people have learned). Hence, it is optimal to keep simple taxes if people learn slowly.

---

23 Recall our distinction of \( S^H_j = S^r : \frac{\partial p^r(p)}{\partial p_j} \) vs \( S^H_{jl} = S^r : \frac{\partial p^r(p,m)}{\partial p_j} \), holding attention constant. Here \( \frac{\partial p^r(p,m)}{\partial p_j} \) might be estimated as a very short run sensitivity.
7 Conclusion

We revisited the main results from the theory of optimal taxation, with behavioral agents. A natural extension is to dynamic settings. One effect can be anticipated. Suppose agents do not pay much attention to the after-tax rate of returns on investment. Then, it may be optimal to tax capital income, which will contradict optimal taxation results under rational agents, which often prescribe zero taxation of capital income. We plan to develop this issue in future work.
References


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


8 Appendix: Notations

$c$: consumption
$h$: index for household type $h$
$m, M$: attention vector, matrix
$p$: pre-tax price
$p^s$: perceived price
$q = p + \tau$: after-tax price
$L$: government’s objective function. $L^W, L^G$ are respectively the consumer welfare and Revenue raising part of the government’s objective function.

$S_j, S^H_j$: Column of the Slutsky matrix when price $j$ changes.

$w$: personal income
$W$: social utility
$\tau$: tax
$\tau^s$: perceived tax
$u(c)$: true utility
$u(c)$: perceived utility
$v(p, w)$: true indirect utility
$\psi_i$: demand elasticity for good $i$
$\lambda, \Lambda = \lambda - 1$: weight on revenue raised in planner’s objective
$\phi$: speed of learning
$\xi$: externality
$\chi$: quantity recommend by the nudge

Multiperson Ramsey
$\gamma^h$ (resp. $\gamma^{\xi,h}$): marginal social utility of income (resp. adjusted for externalities)
$\tau^b$: ideal corrective tax for misperceptions
$T$: tax schedule (e.g. $T = \{\tau_1, \ldots, \tau_n, r_0\}$, where $r_0$ is a lump-sum transfer).
$\theta$: parametrization of perception functions

Nonlinear income tax
$\gamma(z)$: marginal social utility of income
$g(z)$: social welfare weight
$h(z)$ (resp. $h^*(z)$): density (resp. virtual density).

$n$: agent’s wage, also the index of his type
$q(z)$: marginal retention rate, locally perceived
$Q = (q(z))_{z \geq 0}$: vector of marginal retention rate
$r_0$: tax rebate at 0 income
9 Appendix: Behavioral Consumer Theory

Here we develop behavioral consumer theory with nonlinear budget. The agent faces a problem:
\[ \max_c u(c) \text{ s.t. } B(p, c) \leq w. \]

When the budget constraint is linear, \( B(p, c) = p \cdot c, \)

\[ B_{p_j} = c_j, B_{c_j} = p_j \text{ with a linear budget} \]

In general (e.g. in Mirrlees), the budget constraint \( B(p, c) \) could be non-linear.

The agent, whose utility is \( u(c) \), may not completely maximize. Instead, his policy is described by \( c(p, w) \), which exhausts his budget: \( B(c(p, w), p) = w. \)

Though this puts very little structure on the problem, some basic relations can be derived. We then propose specific behavioral models and an application to welfare with inattention.

9.1 Abstract general framework

The indirect utility is defined as \( v(p, w) = u(c(p, w)) \), and the expenditure function as \( e(p, \hat{u}) = \min_c B(c, p) \text{ s.t. } u(c, p) \geq \hat{u}. \)

This implies: \( v(p, e(p, \hat{u})) = \hat{u} \) (with \( \hat{u} \) a real number). Differentiating with respect to \( p_j \), this implies

\[ \frac{v_{p_j}}{v_w} = -e_{p_j}. \] (56)

We define the compensated-demand based Slutsky matrix, defined as:

\[ S_j^{C}(p, w) := c_{p_j}(p, w) + c_w(p, w) B_{p_j}(c, p) |_{c = e(p, w)} \] (57)

The Hicksian demand is: \( h(p, \hat{u}) = c(p, e(p, \hat{u})) \), and the Hicksian-demand based Slutsky matrix is defined as: \( S_j^{H}(p, \hat{u}) := h_{p_j}(p, \hat{u}). \)

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: \( S_j^{C}(p, w) = \partial_x c(p + x, B(c(p, w), p + x)) |_{x=0} \), while \( S_j^{H}(p, w) = \partial_x c(p + x, e(p + x, v(p, w))) |_{x=0} \).
i.e., using (56),
\[ S^H_j(p, w) = c_{pj}(p, w) - c_w(p, w) \frac{v_{pj}(p, w)}{v_w(p, w)} \]  

(58)

In the traditional model, \( S^C = S^H \), but we shall see that this won’t be the case in general. 

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

\[ B_c \cdot c_w = 1, \quad B_c \cdot c_{pi} = -B_{pi}, \quad \frac{u_c}{v_w} c_w = 1 \]  

(59)

The first two come from differentiating \( B(c(p, w), p) = w \). The third one comes from differentiating \( v(p, w) = u(c(p, w)) \) with respect to \( w \), so \( v_w = u_c c_w \).

**Proposition 24** (Roy, abstract) With \( v(p, w) \), we have

\[ \frac{v_{pj}}{v_w} = -B_{pj}(c(p, w), p) + D_j \]  

(60)

where

\[ D_j(p, w) := -\tau^b(p, w) \cdot c_{pj}(p, w) = -\tau^b \cdot S^H_j = -\tau^b \cdot S^C_j \]  

(61)

and the ideal misoptimization corrective tax is defined to be:

\[ \tau^b(p, w) := B_c(c(p, w), p) - \frac{u_c(c(p, w))}{v_w(p, w)} \]  

(62)

When the agent is the traditional rational agent, \( \tau^b = 0 \). In general, \( \tau^b \cdot c_w(p, w) = 0 \).

**Proof**: Relations (59) imply:

\[ \tau^b \cdot c_w = \left( B_c - \frac{u_c}{v_w} \right) c_w = 1 - 1 = 0. \]  

(63)

Next, we differentiate \( v(p, w) = u(c(p, w)) \)

\[ \frac{v_{pi}}{v_w} = \frac{u_c c_{pi}}{v_w} = \frac{(u_c - v_w B_c + v_w B_e) c_{pi}}{v_w} \]

\[ = \frac{(u_c - v_w B_e) c_{pi}}{v_w} - B_{pi} \text{ as } B_c \cdot c_{pi} = -B_{pi} \text{ from (59)} \]

\[ = -\tau^b \cdot c_{pi} - B_{pi} \]  

(64)

---

24 See Aguiar and Serrano (2015) for a recent study of Slutsky matrices with behavioral models.
Next,
\[ D_j = -\tau^b \cdot c_{pj} = -\tau^b \cdot \left( S^H_j + c_w (p, w) \frac{v_{pj}(p, w)}{v_w(p, w)} \right) \text{ by (58)} \]
\[ = -\tau^b \cdot S^H_j \text{ as } \tau^b \cdot c_w = 0 \]  
(65)

Likewise, (57) gives, using again \( \tau^b \cdot c_w = 0 \)

\[ D_j = -\tau^b \cdot c_{pj} = -\tau^b \cdot \left( S^C_j - c_w B_{pj} \right) = -\tau^b \cdot S^C_j \]

\( \Box \)

**Proposition 25 (Slutsky relation modified)** With \( c(p, w) \) we have

\[ c_{pj} = -c_w B_{pj} + S^H_j + c_w D_j = -c_w B_{pj} - c_w \left( \tau^b \cdot S^H_j \right) + S^H_j \]

\[ c_{pj} = -c_w B_{pj} + S^C_j \]

and

\[ S^C_j - S^H_j = c_w D_j = -c_w \left( \tau^b \cdot S^H_j \right) \]

(66)

**Proof.**

\[ c_{pj} = c_w \frac{v_{pj}(p, w)}{v_w(p, w)} + S^H_j \text{ by (58)} \]

\[ = c_w \left( -B_{pj} + D_j \right) + S^H_j \text{ by Proposition 24} \]

Also, (57) gives: \( c_{pj} = -c_w B_{pj} + S^C_j \). \( \Box \)

**Lemma 6** We have

\[ B_c S^C_j = 0, \quad B_c S^H_j = -D_j. \]

(67)

**Proof** Relations (59) imply \( B_c S^C_j = B_c \left( c_{pj} + c_w B_{pj} \right) = -B_{pj} + B_{pj} = 0 \). Also, \( B_c S^H_j = B_c \left( S^C_j - c_w D_j \right) = -D_j. \) \( \Box \)

### 9.2 Application in Specific Behavioral Models

#### 9.2.1 Misperception model

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

\[ c(p, w) = h^r(p^s(p, w), v(p, w)) \]

(68)
where \( h^r(p^*, u) \) is the Hicksian demand of a rational agent would perceived prices \( p^*(p, w) \).

**Proposition 26** Take the misperception model. Then, with \( S^r(p, w) = h_{\nu^*}^r(p^*(p, w), v(p, w)) \) the Slutsky matrix of the underlying rational agent, we have:

\[
S^H_j(p, w) = S^r(p, w) \cdot p^*_j(p, w)
\]

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

\[
\tau^b = B_c(c, p) - \frac{B_c(c, p^s)}{B_c(c, p^s) \cdot c_w}
\]

Given \( B_c(p^s, c) \cdot S^H_j = 0 \), we have:

\[
D_j = (B_c(p, c) - B_c(p^s, c)) \cdot S^H_j = B_c(p, c) \cdot S^H_j
\]

so that

\[
D_j = \tau^b \cdot S^H_j \text{ with } \tau^b := B_c(p, c) - B_c(p^s, c)
\]

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

**Proof** Given \( c(p, w) = h^r(p^s(p, w), v(p, w)) \), we have \( c_w = h^r_u v_w \). Then,

\[
S^H_j = c_{pj}(p, w) - c_w(p, w) \frac{v_{pj}(p, w)}{v_w(p, w)}
\]

\[
= h^r_{\nu^*} p^*_j(p, w) + h^r_u v_{pj} - c_{wj} \frac{v_{pj}}{v_w}
\]

\[
= S^r_{pj}(p, w) + h^r_{\nu^*} v_{pj} - h^r_u v_w \frac{v_{pj}}{v_w} \text{ as } c_w = h^r_u v_w
\]

\[
= S^r_{pj}(p, w)
\]

Next, observe that the demand satisfies \( u_c(p, w) = \Lambda B_c(p^s, c) \) for some Lagrange multiplier \( \Lambda \), and that \( B_c(p^s, c) \cdot S^r = 0 \) for a rational agent (see equation (67) applied to that agent). So, \( B_c(p^s, c) \cdot S^H = 0 \).

\[
-D_j(p, w) = \tau^b \cdot S^H_j = \left( B_c - \frac{u_c}{v_w} \right) \cdot S^r_{pj}(p, w)
\]

\[
= \left( B_c - \frac{\Lambda B_c(p^s, c)}{v_w(p, c)} \right) \cdot S^r_{pj}(p, w)
\]

\[
= B_c \cdot S^r_{pj}(p, w) = (B_c - B_c(p^s, c)) \cdot S^r_{pj}(p, w)
\]

55
Finally, we have \( \frac{u_c}{v_{cw}} = \Lambda B_c (c, p^s) \) for some scalar \( \Lambda > 0 \). Given (59) \( \frac{u_c(c(p,w))}{v_{cw}(p,w)} = \frac{u_c}{u_c-cw} = \frac{B_c(c,p^s)}{B_c(c,p^s)-cw} \) (indeed, both are equal to \( \frac{u_c}{u_c-cw} \)).

We note that \( u_c \cdot S^H = 0 \) in the (static) misperception model (this is because \( u_c = \Lambda B_c (c, p^s) \) for some scalar \( \Lambda \), and \( B_c (c, p^s) \cdot S^H = 0 \) from Proposition 27). This is not true in the decision-utility model.

**Representation lemma**

**Lemma 7** (Representing an abstract demand by a misperception). Given an abstract demand \( c(p,w) \), and a utility function \( u(c) \) we can define the function:

\[
p^s(p,w) := \frac{u_c(c(p,w))}{v_{cw}(p,w)}
\]

(71)

Then, the demand function can be represented as that of a sparse agent with perceived prices \( p^s(p,w) \).

\[
c(p,w) = c^s(p,p^s(p,w),w).
\]

(72)

**Proof.** The demand of a sparse agent \( c^s(p,p^s,w) \) is characterized by \( u_c(c^s(p,p^s,w)) = \lambda p^s \) for some \( \lambda \), and \( p \cdot c = w \). By construction, we have \( u_c(c^s(p,p^s,w)) = \lambda p^s \) for \( p^s = \bar{p}^s(p,w) \). Hence, the representation is valid. [xx talk about potentially multiple solutions. The condition is very mild: given a \( c = c(p,w) \), there's no other \( c' \) with \( p \cdot c' = w \), \( u_c(c') = u_c(c) \), and \( u(c') > u(c) \).]

We note that for any \( p^s(p,w) := ku_c(c(p,w)) \) for some \( k > 0 \), we have \( \frac{u_c(c(p,w))}{v_{cw}(p,w)} = \frac{p^s}{p^s-cw} \) (indeed, both are equal to \( \frac{u_c}{u_c-cw} \)).

**9.2.2 Decision-utility model**

In the decision-utility model there is a “true” utility function \( u(c) \), and a perceived utility function \( u^b(c) \). Demand is \( c(p,w) = \arg \max_c u^b(p,c) \) s.t. \( B(p,c) \leq w \).

Consider the “parallel” agent who’s rational with utility \( u^b \). We call \( u^b(p,w) = u^b(c(p,w)) \) his utility. For that parallel agent, rational agent, call \( S_{j}^{b,r}(p,w) = c_p(p,w) + c_w(p,w)'c \) his Slutsky matrix.

**Proposition 27** In the decision-utility model, \( S_{j}^{C} = S_{j}^{b,r} \) is the Slutsky matrix of a rational agent with utility \( u^b(c) \). The ideal misoptimization corrective tax is:

\[
\tau^b = \frac{u^b_c(c(p,w))}{v_{cw}(p,w)} - \frac{u_c(c(p,w))}{v_{cw}(p,w)}.
\]
9.2.3 Hybrid Model: Agent maximizing the wrong utility function with the wrong prices

Suppose now an agent with true problem \( \max_c u(c) \) s.t. \( B(p,c) \leq w \), but who maximizes \( \max_{c|p^s} u^b(c) \) s.t. \( B(p,c) \leq w \) with both the wrong utility and the wrong prices. This is hybrid of the two previous models.

In terms of decision (if not welfare), the agent is a misperceiving agent with utility \( \varphi^b \) and perceived prices \( \rho^b \). Call \( v^b(p,w) = u^b(c(p,w)) \) and \( h^b(p^s, \hat{\varphi}) = \arg\min_c B(p^s,c) \) s.t. \( u^b(c) \geq \hat{\varphi} \) the indirect utility function (of that misperceiving agent) and the rational compensated demand of that agent with utility \( u^b \). Then, our agent has demand:

\[
\begin{align*}
    c(p,w) &= h_r^b(p^s(p,w), v^b(p,w)) \\
\end{align*}
\]

Proposition 28 (Agent misperceiving both utility and prices) Take the model of an agent maximizing the wrong utility function \( u^b(c) \), with the wrong perceived prices \( \rho^b \). Call \( S_{ij}^r(p,w) = h_r^b(p^s(p,w), v^b(p,w)) \) the Slutsky matrix of the underlying rational agent who has utility \( u^b \), and define

\[
S_{ij}^b(p,w) := S_{ij}^r(p,w) \cdot p^s_j(p,w)
\]

i.e. \( S_{ij}^b = \sum_k S_{ik}^r \frac{\partial p^r_k(p,w)}{\partial \rho_j} \), where \( \frac{\partial p^r_k(p,w)}{\partial \rho_j} \) is the matrix of marginal perception. Then,

\[
\begin{align*}
    S^C_j(p,w) &= S^b_j(p,w) + c_w \left( \frac{v^b_j}{v^b_w} + B_{pj} \right) \\
    S^H_j(p,w) &= S^b_j + c_w \left( \frac{v^b_j}{v^b_w} - \frac{v_{pj}}{v_w} \right) \\
    S^b_j(p,w) &= c_{pj}(p,w) - c_w \frac{v^b_j}{v^b_w} \\
\end{align*}
\]

We can write

\[-D_j = \tau^b \cdot S^b_j \]

with:

\[
\tau^b := (B_c(c,p) - B_c(p^s, c)) + \left( \frac{u^b_c}{u^b_w} - \frac{u^b_c}{v^b_w} \right)
\]

Finally, \( B_c(p^s, c) \cdot S^b_j = 0 \).

This tax \( \tau^b \) is the sum of two gaps: between the prices \( B_c(c,p) - B_c(p^s, c) \) and perceived prices, and between true utility and perceived utility \( \left( \frac{u^b_c}{v^b_w} - \frac{u^b_c}{v^b_w} \right) \).
Proof. So, with $w = \mathbb{P}_{w}^{a}(p, w)$, and use $c(p, w) = h^{r,b}(p^{s}, v^{b}(p, w))$:

$$c_{w}(p, w) = h^{r,b}_{w}v^{b}_{w}$$
$$c_{pj}(p, w) = h^{r,b}_{pj} + h^{r,b}_{u}v^{b}_{j}$$
$$= S^{b}_{j} + c_{w}v^{b}_{w}$$

Using (57) and (58) gives:

$$S^{C}_{j} = c_{pj}(p, w) + c_{w}(p, w) B_{pj}(c, p) = S^{b}_{j} + c_{w} \left( \frac{v^{b}_{j}}{v^{w}_{w}} + B_{pj} \right)$$
$$S^{H}_{j} = c_{pj}(p, w) - c_{w}(p, w) \frac{v_{pj}(p, w)}{v_{w}(p, w)} = S^{b}_{j} + c_{w} \left( \frac{v^{b}_{j}}{v^{w}_{w}} - \frac{v_{pj}(p, w)}{v_{w}(p, w)} \right)$$

We have

$$B_{c}(p^{s}, c) \cdot S^{b}_{j} = B_{c}(p^{s}, c) \cdot h^{r}(p^{s}, v^{b}) \cdot M_{j} = 0 \text{ as } B_{c}(p^{s}, c) \cdot h^{r}(p^{s}, v^{b}) = 0$$

Recall also that $\frac{w^{b}_{j}}{v^{w}_{w}} = \Lambda B_{c}(p^{s}, c)$ as the agent maximizes with perceived prices $p^{s}$. Hence,

$$-D_{j} = \left( B_{c}(c, p) - \frac{u_{c}}{v_{w}} \right) S^{b}_{j}$$
$$= \left( (B_{c}(c, p) - B_{c}(p^{s}, c)) - \left( \frac{u_{c}}{v_{w}} - \Lambda B_{c}(p^{s}, c) \right) \right) S^{b}_{j} \text{ as } B_{c}(p^{s}, c) \cdot S^{b}_{j} = 0$$
$$= \left( B_{c}(c, p) - B_{c}(p^{s}, c) - \left( \frac{u_{c}}{v_{w}} - \frac{u^{b}_{j}}{v^{b}_{w}} \right) \right) \cdot S^{b}_{j}$$
$$= \tau_{b} \cdot S^{H}_{j}$$

\[\square\]

9.3 Attention as a good

9.3.1 Interpreting attention as a good

We propose a simple abstract way to think about attention, including its potentially suboptimal allocation. Call $c = (C, m)$ the “metagood” made of both regular commodities $C$ and attention $m$, which are both vectors. Utility is $u(c) = u(C, m)$. The framework applies $u(c)$. There is a demand $c(p, w)$. We can think that attention has market price 0 (it could have a non-zero price, for instance if $m$ is the amount of computer power one uses to optimize).

The concrete example of misperception framework is worth keeping in mind for concreteness.
We have the demand \( C^*(p, p^*, w, m) = \arg \max_{C\mid p^*} u(C, m) \) s.t. \( p \cdot (C, m) \leq w \), which depends on perceived prices \( p^* \). Then, given \( p^* (p, w, m) \), demand is \( C(p, w, m) = C^*(p, p^* (p, m, w), w, m) \). If \( m \) is optimally allocated, \( m(p, w) = \arg \max_m u(C(p, w, m), m) \). In general (even with non-optimal attention), given an attention policy \( m(p, w) \), demand is \( C(p) = C(p, w, m(p, w)) \).

**Characterizing optimal allocation of attention** Suppose we have a constraint: \( c(p, w) = c(p, w, \theta) \) for some parameter \( \theta \). For instance, suppose that \( c(p, w, \theta) = (C(p, w, m(\theta)), m(\theta)) \); when \( m(\theta) = \theta \), we’re considering the potentially optimal allocation of attention, as attention affects directly the choice of goods. If \( m = (m_1, m_2, m_3) = (\theta_1, \theta_2, \theta_3) \), we captures that the attention to goods 2 and 3 have to be the same.\(^{25}\)

**Proposition 29** (Characterizing optimal allocation of attention) The first order condition for the optimal allocation of parameter \( \theta \) (i.e., \( \theta(p, w) = \arg \max_\theta u(c(p, w, \theta)) \)) is:

\[
\tau^h \cdot c_\theta (p, w, \theta) = 0
\]  

**(Proof)** The FOC is \( u_c c_\theta = 0 \). We note that \( B_c \cdot c_\theta = 0 \) by budget constraint: \( B(c(p, w, \theta)) = w \). So,

\[
\tau^b \cdot c_\theta = \left( \frac{u_c(c, p)}{v_w(p, w)} - B_c \right) \cdot c_\theta = \frac{u_c(c, p) \cdot c_\theta}{v_w(p, w)}
\]

so that \( \tau^b \cdot c_\theta = 0 \) iff \( u_c \cdot c_\theta = 0 \). \( \square \)

**Proposition 30** (Value of \( D_j \) when attention is optimal). When attention is of the form \( c(p, w, \theta) = (C(p, w, m(\theta)), m(\theta)) \), and is optimally chosen, then

\[
-D_j = \tau^h_C \cdot C_{p_j} (p, w, m)|_{m=m(\theta(p, w))} = \tau^h_C \cdot S^H_{j|m} (p, w, m)|_{m=m(\theta(p, w))} = \tau^h_C \cdot S^C_{j|m} (p, w, m)|_{m=m(\theta(p, w))}
\]

where \( \tau^h_C = B_C(C, p) - \frac{u_c(C, m)}{v_w(p, w)} \) is the ideal tax restricted to goods consumption, and \( S^H_{j|m} \) and \( S^C_{j|m} \) are the Slutsky matrices \( S^H \) and \( S^C \) holding attention constant, i.e. associated to decision \( C(p, w, m) \) with constant \( m = m(\theta(p, w)) \).

\(^{25}\)In a model of noisy decision-making à la Sims (2003), the same logic exactly applies, except that all quantities are stochastic. The consumption is a random variable \( c(p, w, \tilde{\omega}) \), where \( \tilde{\omega} \) indexes noise, rather than a deterministic function. Then, utility is \( U(c(p, w)) := E[u(c(p, w, \tilde{\omega}))] \), \( S^H(p, w) \) is likewise a random variable. We do not pursue this framework further here, at it is hard to solve beyond linear-quadratic settings, e.g. with Gaussian distribution of prices – which in turn generates potentially negative prices.
Proof We have

\[-D_j = \tau^b \cdot c_{p_j} (p, w, \theta) = \tau^b \cdot \left[ (C_{p_j} (p, w, m), 0) + c_{\theta} (p, w, \theta) \theta_{p_j} (p, w) \right] \]

\[= \tau^b \cdot (C_{p_j} (p, w, m), 0) \text{ as } \tau^b \cdot c_{\theta} = 0 \]

\[= (\tau^b_c, \tau^b_m) \cdot (C_{p_j} (p, w, m), 0) \]

\[= \tau^b_c \cdot C_{p_j} (p, w, m) \cdot \]

\[= \tau^b_c \cdot S_j^{C, m} = \tau^b_c \cdot S_j^{H, m} \]

where \(\tau^b_c \cdot S_j^{H, \text{constant} m}\) is the Slutsky matrix with a constant \(m\). □

When the attention is costless and automatic Another benchmark is the “automatic attention whose cost are not accounted for”. Suppose that attention \(m\) just moves with prices, but as an automatic process whose “cost” is not counted: this is, \(u (C, m) = u (C)\) and attention has 0 price, \(p_m = 0\). This is the way it is often done in behavioral economic (see however Bernheim and Rangel 2009): people choose using heuristics, but the “cognitive cost” associated with a decision procedure isn’t taken into account in the agent’s welfare (largely, because it is very hard to measure, and that revealed preference techniques do not apply).

Proposition 31 (Value of \(D_j\) when attention is fixed, or costless and automatic). When attention is either fixed, or costless but automatic, then

\[-D_j = (\tau^b_c, 0) \cdot S_j^H (p, w) = \sum_{i=1}^{n} \tau^b_{C_i} S_{ij}^H \]

\[= (\tau^b_c, 0) \cdot S_j^C (p, w) = (\tau^b_c, 0) \cdot c_j (p, w) \]

This is, only the components of \(\tau^b\) and the Slutsky matrix linked to commodities matter.

Proof

We have \(\tau^b = (\tau^b_c, \tau^b_m) = (\tau^b_c, 0)\) as \(u_m = 0\).

\[-D_j = \tau^b \cdot S_j^H (p, w) = \tau^b_c \cdot S_{C, j}^H. \]

□

Misperception example In the misperception model with attention policy \(m (p, w)\), we have:

\[c (p, w) = (C^n [p, p^* (p, w, m (p, w)), v (p, w)], m (p, w)) \]

When attention is optimally chosen, we can apply Proposition 30 with \(m (\theta) = \theta\). This gives:
\[-D_j = \tau_b^c \cdot S^{H,m}_{Cj} \text{ with} \]

\[S^{H}_{Cjm} = S^r p^*_p (p, w, m) \tag{77}\]

i.e. the Slutsky matrix has the the sensitivity with fixed attention. Hence, we have both \(-D_j = \tau_c^b \cdot S^{H,m}_{j} \) when attention is optimal.

When attention is not necessarily optimal, we also have (from (70)), using again decomposition \(\tau^b = (\tau^b_c, \tau^b_m) \):

\[-D_j = \tau^b \cdot S = \tau^b_c \cdot S^H_{Cj} + \tau^b_m \frac{\partial m}{\partial p_j} \]

where \(S^H_{Cj} = S^r \cdot p^*_p (p, w)\), where now the total derivative matters, including the variable attention.

When attention is costless, \(\tau^b_m = 0\) and

\[-D_j = \tau^b_c \cdot S^H_{Cj} \]

### 9.3.2 Attention as a good: examples

Take decision utility have \(u^b(c_0, c_1, m) = c_0 + U(c_1) - q(m)\), with \(U(c) = \frac{ac - \frac{1}{2}c^2}{\Psi}\) and attention technology \(p^*_1(p_1, m) = p^d_1 + m\tau_1\), where \(\tau_1\) is a tax (implicitly, that assume \(\tau^d_1 = 0\)). Full utility is \(u(c_0, c_1, m) = c_0 + U(c_1) - Ag(m)\), where \(A = 0\) in the “no attention in welfare” case, and \(A = 1\) in the “optimally allocated attention” case.

We assume \(p_0 = 1, \Psi > 0\). Given attention \(m\), demand satisfies \(U'(c_1) = p^s\), so \(c^*_1(p^s) = a - \Psi p^s\). The perceived tax is:

\[\tau^s_1 = m(\tau_1) \tau_1\]

and demand is

\[c_1 = a - \Psi (p^d_1 + m(\tau_1) \tau_1)\]

The losses from inattention are \(\frac{1}{2}u_{c_1 c_1} (c^*_1 - c_1)^2 = -\frac{1}{2} \Psi \tau^2 (1 - m)^2\). (This is always true to the leading order, and here this is exact as the function is quadratic; see Gabaix 2014, Lemma 2). Hence, the optimal attention problem is:

\[m(\tau_1) = \arg \max_m -\frac{1}{2} \Psi \tau^2 (1 - m)^2 - g(m) ,\]

whose first order condition is:

\[g'(m) = \Psi (1 - m) \tau^2_1 \tag{78}\]

The Slutsky matrix with constant \(m\) has:

\[S^H_{11|m} = \frac{\partial c_1 (p_1, m)}{\partial p_1} = -\Psi m\]
while with variable attention $m(p)$, we have:

$$S_{11}^H = \frac{dc_1(p_1, m(p_1))}{dp_1} = -\Psi (m + \tau m'(\tau_1))$$

$$= -\Psi (m + \tau \Psi \tau_1 (1 - m)^2)$$

$$S_{21}^H = \frac{\partial m}{\partial p_1} = m'(\tau_1)$$

Then, we have: $\tau^b = (0, q - q^a, Ag'(m)) = (0, \tau_1 (1 - m), Ag'(m))$, and given $\tau = (0, \tau_1, 0)$, so

$$\tau - \tau^b = (0, \tau_1 m, -Ag'(m))$$

$$S_1^H = (0, -\Psi (m + \tau_1 m'(p)), m'(p_1))$$

Applying Proposition 8 gives:

$$\frac{\partial L(\tau, w)}{\partial \tau_1} = (\lambda - \gamma)c_1 + \lambda (\tau - \tau^b) \cdot S_1^H$$

$$= (\lambda - \gamma)c_1 - \Psi \tau_1 m (m + \tau_1 m'(p)) - Ag'(m) m'(p_1)$$

Normalize $\lambda = 1, \gamma = 1 - \Lambda$, and define $\psi_1(c_1) = \Psi/c_1$. First, when $m_1$ is exogenous, we verify our formula from section 2

$$\frac{\partial L(\tau, w)}{\partial \tau_1} = \Lambda c_1 - \Psi m^s \tau_1$$

i.e.

$$\tau^s_1 = \frac{\Lambda}{m \psi_1}$$

$$\tau_1 = \frac{\Lambda}{m^2 \psi_1}$$

Next, in the "no attention in welfare" case, $A = 0$

$$\frac{\partial L(\tau, w)}{\partial \tau_1} = \Lambda c_1 - \Psi \tau^s_1 \frac{d \tau^s_1 (\tau_1, m_1(\tau_1))}{d \tau_1} = \Lambda c_1 - \Psi \tau^s_1 (m_1 + \tau_1 m'(\tau_1))$$

so

$$\tau^s_1 = \frac{-c_1 \Lambda}{S_{11}^H} = \frac{\Lambda}{(m + \tau_1 m'(\tau_1)) \psi_1} \leq \frac{\Lambda}{m \psi_1}$$

$$\tau_1 = \frac{\Lambda}{(m^2 + \tau_1 mm'(\tau_1)) \psi_1}$$
Finally, in the “optimally allocated attention” case, $A = 1$. First, we verify:

$$-D_1 = \tau^b \cdot S^H = (0, \tau_1 (1 - m), g' (m)) \cdot (0, -\Psi (m + \tau_1 m' (p)), m' (p_1))$$

$$= -\Psi (m + \tau_1 m' (p)) \tau_1 (1 - m) + g' (m) m' (p_1)$$

$$= -\Psi m \tau_1 (1 - m)$$

$$= - (\tau_1 - \tau^*_1) \Psi m$$

$$= \tau^b \cdot S^H_{jm} (p, w, m)$$

with $\tau^b = \tau_1 - \tau^*_1 = (1 - m) \tau_1$ and $S^H_{jm} (p, w, m) = -\Psi m$.

$$\frac{\partial L (\tau, w)}{\partial \tau_1} - \Lambda c_1 = (\tau - \tau^b) \cdot S^H_1 = \tau \cdot S^H_1 - \tau^b \cdot S^H_1 = \tau \cdot S^H_1 - \tau^b \cdot S^H_1_{|m}$$

$$= -\Psi \tau (m + \tau_1 m' (\tau_1)) + \Psi \tau (1 - m) m$$

$$= -\Psi \tau (m^2 + \tau m' (\tau))$$

$$\frac{\partial L (\tau, w)}{\partial \tau_1} = \Lambda c_1 - \Psi \tau (m^2 + \tau m' (\tau))$$

so

$$\tau = \frac{\Lambda}{(m (\tau)^2 + \tau m' (\tau)) \psi_1}.$$

We see the following.

**Proposition 32** In the interior region where attention has an increasing cost ($\tau m (\tau) m' (\tau) > 0$), we have:

$$\tau^{Optimal attention}_1 < \tau^{No cost of attention}_1 < \tau^{Fixed attention}_1.$$  \hfill (79)

### 9.3.3 Optimal tax with endogenous, optimally chosen attention

There is just one taxed good. The case with many, independent taxed goods follows. We assume $\tau^d = 0$.

The consumer chooses:

$$m (\tau) = \arg \min \left\{ -\frac{1}{2} \psi y \tau^2 (1 - m)^2 - \kappa g^1 (m) \right\}$$

and the planner chooses: $\tau (\Lambda) := \arg \max_{\tau} L (\tau, \Lambda)$

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

**A lemma on scaling** We show that it is enough to compute the solution in the case $\psi = y = \kappa = 1$. 

63
Lemma 8 Suppose that when \( \psi = y = \kappa = 1 \) the optimal tax is \( \tau' = f(\Lambda) \) and optimal attention is \( m^1(\tau') \). Then, in the general case it is:

\[
\tau(\Lambda) = \sqrt{\frac{\kappa}{\psi y}} f \left( \Lambda \sqrt{\frac{y}{\kappa \psi}} \right)
\]

and the attention is \( m(\tau) = m^1 \left( \tau \sqrt{\frac{\psi y}{\kappa}} \right) \).

For instance, in the basic rational case, \( f(\Lambda) = \Lambda \) and \( m^1(\tau') = 1 \).

Proof This is a simple scaling argument. We define

\[
m^1(\tau') := \arg\min \frac{-1}{2}(1-m)^2 - g^1(m)
\]

\[
L^1(\tau', \Lambda') = \frac{1}{2} m^1(\tau')^2 - g^1(m^1) + \Lambda' \tau
\]

\[
\tau' := \tau \sqrt{\frac{\psi y}{\kappa}}
\]

\[
\Lambda' := \Lambda \sqrt{\frac{y}{\kappa \psi}} = \frac{\Lambda y}{\kappa} \tau'
\]

Then, we have:

\[
m(\tau) = \arg\min \frac{-1}{2} \frac{\psi y}{\kappa} \tau^2 (1-m)^2 - g^1(m)
\]

\[
= m^1(\tau')
\]

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

\[
= \kappa \left[ -\frac{1}{2} \frac{\psi y}{\kappa} \tau^2 m(\tau)^2 - g^1(m) + \frac{\Lambda y}{\kappa} \tau \right]
\]

\[
= \kappa L^1(\tau', \Lambda')
\]

Hence, as \( \tau' = f(\Lambda') \) at the optimum. \( \square \)

Example with continuously adjusting attention We have \( g(m) = -\kappa \ln (1-m) \), so that attention is \( m(\tau) = \left( 1 - \frac{1}{\sqrt{\psi y} \tau} \right)_+ \). Indeed, \( \arg\min \frac{\alpha^2}{2} (1-m)^2 + g(m) \) is \( m = (1 - \frac{1}{\sigma})_+ \).

Proposition 33 In the above setup with optimal attention, the optimal tax is \( \tau_i = \sqrt{\frac{\kappa}{\psi y_i}} f \left( \Lambda \sqrt{\frac{y_i}{\kappa \psi_i}} \right) \),
for the continuous function

\[ f(\Lambda) = \frac{\Lambda + 1 + \sqrt{(\Lambda + 1)^2 - 4}}{2} \quad \text{for } \Lambda \geq 1 \]

\[ = 1 \quad \text{for } \Lambda < 1. \]

Also, \( m^1(\tau') = (1 - \frac{1}{\tau'})_+ \).

**Proof.** We first reason in the case \( \psi = y = \kappa = 1 \). Then, \( m(\tau) = (1 - \frac{1}{\tau})_+ \) and

\[ L(\tau) = -\frac{1}{2}m(\tau)^2 \tau^2 - g(m) + \Lambda \tau \]

Then, for \( \tau > 1 \),

\[ L'(\tau) = 1 - \frac{1}{\tau} - \tau + \Lambda \]

so \( \tau \) is the greater root of:

\[ \tau + \frac{1}{\tau} = \Lambda + 1 \]

which exists provided \( \Lambda \geq 1 \), i.e.:

\[ f(\Lambda) = \frac{\Lambda + 1 + \sqrt{(\Lambda + 1)^2 - 4}}{2} \quad \text{for } \Lambda \geq 1 \]

\[ = 1 \quad \text{for } \Lambda < 1. \]

\[ \square \]

**An example with 0-1 attention** A concrete example of attention choice is: \( m(\tau) = \arg\max_m -\frac{1}{2} \psi \tau^2 (1 - g(m)) \) with \( g(m) = \frac{1}{2} \kappa^2 \left[ 1 - (1 - m)^2 \right] \). Then, the solution is

\[ m(\tau) = 1_{\tau > \tau_*}, \quad \tau_* := \frac{\kappa}{\sqrt{\psi}} \quad (80) \]

As an aside, a fixed cost \( g(m) = \frac{\kappa^2}{2} m_{m>0} \) gives the same result.

**Proposition 34** The optimal tax is \( \tau_i = \sqrt{\frac{\kappa}{\psi \psi_i h}} f\left( \Lambda \sqrt{\frac{\psi}{\psi \psi_i h}} \right) \), for \( f(\Lambda) = 1 \) if \( \Lambda \leq \sqrt{2} + 1 \) and \( f(\Lambda) = \Lambda \) if \( \Lambda > \sqrt{2} + 1 \). Also, \( m^1(\tau') = 1_{\tau' > 1} \).

In that case, the optimal tax has a discontinuity. When \( \Lambda \) is low enough, the planner keeps the taxes at \( \tau_i = \sqrt{\frac{\kappa}{\psi \psi_i h}} \), just below the “detectability threshold” and agents do not pay attention to the tax.
We start with the case $\psi = y = \kappa = 1$. Then, $m(\tau) = 1_{\tau>1}$. For $\tau \leq 1$, $L(\tau) = \Lambda \tau$, so the optimum for $\tau \in [0, 1]$ is $\tau = 1$.

$$L(1) = \Lambda$$

For $\tau > \tau^*$, $m(\tau) = 1$, so $L(\tau) = -\frac{1}{2}\tau^2 + g(1) + \Lambda \tau$, and the optimum is $\tau = \frac{\Lambda}{\psi}$. We have

$$L(\tau) = \frac{\Lambda^2 - 1}{2}$$

So $L(\tau) > L(\tau^*)$ iff $\frac{\Lambda^2 - 1}{2} > \Lambda$, i.e. iff $\frac{\Lambda^2 - 1}{2} > \Lambda$, i.e. iff $\Lambda > \sqrt{2} + 1$.□

## 10 Appendix: Proofs

The online appendix contains further proofs.

**Proof of Lemma 1** Because utility is quasilinear, $v(p, p^s, w) = w + v(p, p^s, 0)$, so $e(p, p^s, 0) = -v(p, p^s, 0)$. We have

$$L(t) = u(c) + \lambda t \cdot c$$

$$= v(p + \tau, p + \tau^s, w) + \lambda t \cdot c(p + \tau, p + \tau^s, w)$$

$$= w - e(p + \tau, p + \tau^s, 0) + (1 + \Lambda) t \cdot c(p + \tau, p + \tau^s, w)$$

By Taylor expansion, around $(\tau, \tau^s) = (0, 0)$, using Propositions 35 and 39, we have:

$$e(p + \tau, p + \tau^s, 0) - e(p, p) = [e_p \tau + e_{p^s} \tau^s] + \left[\frac{1}{2} \tau e_{p^p} \tau + \tau e_{p^p} \tau^s + \frac{1}{2} \tau^s e_{p^p} \tau^s\right] + o(||\tau||^2)$$

$$= [c^d \tau + 0 \cdot \tau^s] + \left[0 + \tau S^r \tau^s - \frac{1}{2} \tau^s S^r \tau^s\right] + o(||\tau||^2)$$

$$= c^d \tau + \tau S^r \tau^s - \frac{1}{2} \tau^s S^r \tau^s + o(||\tau||^2)$$

Using Proposition 37,

$$t \cdot c(p + \tau, p + \tau^s, w) = t \cdot (c^d + c_p \tau + c_{p^s} \tau^s)$$

$$= \tau \cdot (c^d + 0 + S^r \tau^s + o(||\tau||)) = \tau c^d + \tau S^r \tau^s + o(||\tau||^2)$$
\[ L(\tau) - L(0) = -[e(p + \tau, p + \tau^s, 0) - e(p, p)] + (1 + \Lambda) \cdot c(p + \tau, p + \tau^s, w) \]
\[ = -c^d \tau - \tau S^r \tau^s + \frac{1}{2} \tau S^r S^r \tau^s + (1 + \Lambda) \left( \tau c^d + \tau S^r \tau^s \right) + o \left( \|\tau\|^2 \right) \]
\[ = \Lambda \left( \tau c^d + \tau S^r \tau^s \right) - \frac{1}{2} \tau S^r S^r \tau^s + o \left( \|\tau\|^2 \right) \]
\[ = \Lambda \tau c^d - \frac{1}{2} \tau S^r S^r \tau^s + o \left( \|\tau\|^2 \right) + O \left( \|\tau\|^2 \Lambda \right) \]

**Proof of Proposition 2**  Consider first the case \( \alpha = 1, \beta = 0 \). We always have \( \tau^s_i \leq \tau_{\text{max}} \).

Consider the marginal impact of increasing all taxes by \( \varepsilon \). All \( \tau_i \) and \( \tau^s_i \) increase by \( \varepsilon \), and using \( i = (1, ..., 1)' \),

\[ 0 = \frac{dL(\tau + \varepsilon \cdot i)}{d\varepsilon} = \sum_i -\tau^s_i \psi_i y_i + \Lambda \sum_i y_i \geq \sum_i -\tau_{\text{max}} \psi_i y_i + \Lambda \sum_i y_i \]

so \( \tau_{\text{max}} \geq \tau^* := \frac{\Lambda}{\Psi} \). Also, if the tax is uniform, it should be at the rate \( \tau^* = \frac{\Lambda}{\Psi} \).

Suppose by contradiction there is a good \( i \) such that \( \tau_i < \tau_{\text{max}} \). Then, (as \( \frac{\partial r^d(\tau)}{\partial \tau_i} = 0 \) locally), \( \tau^s_i = \frac{\Lambda}{\Psi m_i} > \frac{\Lambda}{\Psi} = \tau^* \) by the assumption in the Proposition. Also, if a good has \( \tau_i = \tau_{\text{max}} \), it satisfies \( \tau^s_i \geq \tau^* \) as \( \tau_{\text{max}} \geq \tau^* \). Hence, for all goods, \( \tau^s_i \geq \tau^* \), and \( \tau_i > \tau^* \) for at least one good. Now, revisit:

\[ 0 = \frac{dL(\tau)}{d\varepsilon} = \sum_i -\tau^s_i \psi_i y_i + \Lambda \sum_i y_i < \sum_i -\tau^* \psi_i y_i + \Lambda \sum_i y_i = 0. \]

We have reached the desired contradiction.

**In the case \( \beta \) and \( m \) sufficiently small.** Loosely speaking the reasoning holds by continuity. A rigorous proof is in Proposition 16 and Prop 44 of the online appendix. \( \square \)

**Proof of Proposition 5**

At the optimum, \( U^{ht}(c^h) = p + \xi \). We suppose \( \tau^d = 0 \) for simplicity. If the agent perceives only \( m^h \tau \), his demand is off the ideal \( c^h \) (up to second order terms) as:

\[ c^h = c^h - \Psi \left( m^h \tau - \xi^h \right) \]

This expression is exact in the quadratic functional form about, and otherwise the leading term of a Taylor expansion of a general function, with now the interpretation \( \Psi = \psi^h c^h \) then. So the welfare loss is:

\[ L^h = W^h - W^h* = \frac{1}{2} m^h \cdot \left( -\Psi \cdot (m \tau - \xi^h) \right)^2 = -\frac{1}{2} \Psi \left( m^h \tau - \xi^h \right)^2 \]

and social welfare is \( L = \sum_h L^h = -\frac{\Psi}{2} \sum_h (m^h \tau - \xi^h)^2 \).

Because \( L_r = -\Psi \sum_h m^h \left( m^h \tau - \xi^h \right) \), the optimal tax is

\[ \tau^* = \frac{\sum_h \xi_h m_h}{\sum_h m_h^2} = \frac{E [\xi_h m_h]}{E [m_h^2]} . \]
Let us calculate \( V = E \left[ (m_h \tau - \xi_h)^2 \right] \) at this optimum \( \tau = \tau^* \),

\[
V = E \left[ m_h^2 \right] \tau^{*2} - 2E \left[ m_h \xi_h \right] \tau^* + E \left[ \xi_h^2 \right] \\
= E \left[ m_h^2 \right] \frac{E \left[ \xi_h m_h \right]^2}{E \left[ m_h^2 \right]} - 2E \left[ m_h \xi_h \right] \frac{E \left[ \xi_h m_h \right]}{E \left[ m_h^2 \right]} + E \left[ \xi_h^2 \right] = - \frac{E \left[ \xi_h m_h \right]^2}{E \left[ m_h^2 \right]} + E \left[ \xi_h^2 \right]
\]

hence the welfare loss is: \( L = - \frac{1}{2} H \Psi \frac{E \left[ \xi_h^2 \right]}{E \left[ m_h^2 \right]} \).

If there is no tax, the loss is (from equation 15):

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

So, \( L = L^{\text{no tax}} \frac{E \left[ \xi_h^2 \right]}{E \left[ m_h^2 \right]} \).

\[\square\]

**Proof of Proposition 6**  
Welfare is \( \sum_h \left[ U^h \left( c^* \right) - \left( p + \xi^h \right) c^* \right] \). The optimal quantity restriction \( c^* \) is characterized:

\[
\frac{1}{H} \sum_h U^{\text{opt}} \left( c^* \right) = p + \frac{1}{H} \sum_h \xi_h
\]

(81)

The welfare loss compared to the first best, which entails \( U^{\text{opt}} \left( c^*_h \right) = p + \xi^h \) is

\[
L^h = \frac{1}{2} U^{\text{opt}} \left( c^*_h - c^* \right)^2 = - \frac{1}{2} \Psi \left( c^*_h - c^* \right)^2
\]

The best consumption satisfies: \( L^Q = \frac{1}{2} \sum_h \left( c^*_h - c^* \right)^2 = 0 \), i.e. \( c^* = E \left[ c^*_h \right] \)

The loss is:

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

\[\square\]

**Proof of Proposition 9**  
We observe that a tax \( \tau_i \) modifies the externality as:

\[
\frac{d \xi}{d \tau_i} = \sum_h \xi_{i,h} \left( \frac{c^h_{i,0} \left( q, w, \xi \right) + c^h_{i} \frac{d \xi}{d \tau_i}}{\frac{d \xi}{d \tau_i}} \right)
\]

so \( \frac{d \xi}{d \tau_i} = \frac{\sum_h \xi_{i,h} c^h_{i,0} \left( q, w, \xi \right)}{1 - \sum_h \xi_{i,h} c^h_{i}} \). The term \( \frac{1}{1 - \sum_h \xi_{i,h} c^h_{i}} \) represents the “multiplier” effect of one unit of pollution on consumption, then on more pollution. So, calling \( \frac{\partial L^{\text{no tax}}}{\partial \tau_i} \) the value of \( \frac{\partial L}{\partial \tau_i} \) without the externality.
(that was derived in Proposition 8)

\[
\frac{\partial L}{\partial \tau_i} - \frac{\partial L^{\text{no } \xi}}{\partial \tau_i} = \frac{d \xi}{d \tau_i} \left( \sum_h W_i^h v^h_{w} v^h_{w} \frac{v^h_{w}}{v^h_{w}} + \lambda \sum_h \tau \cdot c^h (q, w, \xi) \right) = \frac{d \xi}{d \tau_i} \sum_h \left[ \beta^h \frac{v^h_{w}}{v^h_{w}} + \lambda \tau \cdot c^h \right] = \sum_h \xi c^h c^h_i
\]

Using Proposition 8,

\[
\frac{\partial L}{\partial \tau_i} = \sum_h \left( \lambda - \gamma^h \right) c^h_i + \gamma^h D^h + \lambda \tau \cdot S^H_i + \Xi c^h \cdot \left[ c^h \left( -c^h + D^h \right) + S^H_i \right]
\]

\[
= \sum_h \left( \lambda - \gamma^h - \Xi c^h \right) c^h_i + \left( \gamma^h + \Xi c^h \right) D^h + \lambda \left( \tau + \Xi c^h \right) \cdot S^H_i
\]

\[
\frac{\partial L}{\partial \tau_i} = \sum_h \left( \lambda - \gamma^h \xi \right) c^h_i + \gamma^h \xi D^h + \lambda \left( \tau + \Xi \xi \right) \cdot S^H_i
\]

where we defined the externality-augmented social marginal product of income \( \gamma^h \xi := \gamma^h + \Xi c^h \).

\( \square \)

**Proof of Proposition 10**

\[
\frac{\partial L}{\partial \chi} = \sum_h \left[ W_i^h v^h_{w} v^h_{w} \frac{v^h_{w}}{v^h_{w}} + \lambda \left( \tau - \tau^h \right) \right] c^h_i = \sum_h \left[ \beta^h \left( \frac{v^h_{w}}{v^h_{w}} - q + q \right) + \lambda \left( \tau - \tau^h \right) \right] c^h_i
\]

\[
= \sum_h \left[ \beta^h \left( \frac{v^h_{w}}{v^h_{w}} - q \right) + \lambda \tau \right] c^h_i \text{ as } q \cdot c \left( q, w, \chi \right) = w \text{ implies } q \cdot c = 0
\]

In the misperception model with exogenous attention, Proposition 9 gives \( \frac{\partial L(\tau, \tau^s, w)}{\partial \tau^s} = \left[ \gamma^h \left( \tau^s - \tau \right) + \lambda \left( \tau - \tau^h \right) \right] S^h \). So

\[
\frac{\partial L}{\partial \chi} \left( \tau, w, \chi \right) = \sum_h \left[ \lambda \left( \tau - \tau^h \right) + \gamma^h \left( \tau^h - \tau \right) \right] S^h \frac{\partial p^s}{\partial \chi}
\]

\( \square \)

**Proof of Proposition 12** We derived: \( \tau = \left[ M^h S^s M^h \right]^{-1} E \left[ M^h S^r \right] \left( \xi, 0 \right)' \). We have \( \left( E \left[ M^h S^s M^h \right] \right)_{ij} = S^h_{ij} E \left[ m^h_i m^h_j \right] \) and \( \left( E \left[ M^h S^r \right] \right)_{ij} = E \left[ m^h_i \right] S^h_{ij} \). Matrix inversion gives:

\[
\tau_2 = \frac{S^h_{11} S^h_{12} E \left[ m_1 \right] \left( E \left[ m_1^2 \right] E \left[ m_2 \right] - E \left[ m_1 m_2 \right] E \left[ m_1 \right] \right)}{\det E \left[ M^h S^r M^h \right]} \xi
\]

69
Because $E \left[ M^h S^r M^h \right]$ is a dimension $2 \times 2$ and has negative roots (there is a good 0, so that $S^r$ is the block matrix excluding good 0, and has only negative root), $\det E \left[ M^h S^r M^h \right] > 0$. The condition in the Proposition is that $E \left[ m_1^2 \right] E \left[ m_2 \right] - E \left[ m_1 m_2 \right] E \left[ m_1 \right] > 0$. Hence, $\text{sign} \left( \tau_2 \right) = -\text{sign} \left( S_{12} \right)$.

**Proof of Proposition 13** We calculate first the agent’s demand: (given $c^h = b, c^i_1$) (maximizing his subjective value function, not his objective one):

$$c^i_1 = a - \left[ p_1 + \tau_1 + b^h \left( m_2^h p_2 + m_3^h \tau_2 \right) \right],$$

The internality is:

$$\rho^{b,h} = \left( v_c^{b,h} - v_c^{b,h}, v_c^{b,h} - v_c^{b,h} \right) = \left( 0, (1 - m_2^h) p_2 + (1 - m_3^h) \tau_2 \right)$$

and the externality is $\rho^{b,\xi,h} = (0, \xi_s)$. So, as $\tau = (\tau_1, \tau_2)$,

$$\tau - \rho^{b,h} - \rho^{b,\xi,h} = (\tau_1, m_3^h \tau_2 - (1 - m_2^h) p_2 - \xi_s)$$

The behavioral Slutsky matrix is then: $S^{h,h} = \left( \frac{\partial c^i_1}{\partial \tau} \right)_{i,j=1,2} = - \left( \begin{array}{cc} 1 & \psi_h \\ \frac{b_h}{b_h} & \frac{b_h \psi_h}{b_h} \end{array} \right)$.

We apply Proposition 8 (without externality) or its extension in Proposition 9 (with externalities). At the optimum we should have $L_\tau = 0$ with:

$$L_\tau = \sum_h \left( \tau - \rho^{b,h} - \rho^{b,\xi,h} \right) \cdot S^{h,h} = \sum_h - \left( \tau_1 + \psi_h \tau_2 - X_h, \psi_h \tau_1 + \mu_h^2 \tau_2 - \psi_h X_h \right)$$

i.e., using $E \left[ \cdot \right]$ to average over agents:

$$\tau_1 + E \left[ \psi_h \right] \tau_2 = E \left[ X_h \right]$$

$$E \left[ \psi_h \right] \tau_1 + E \left[ \mu_h^2 \right] \tau_2 = E \left[ \psi_h X_h \right]$$

This leads to $\tau_1 = E \left[ x_h \right], \tau_2 = \beta X \psi$.

**Proof of Proposition 16** We call $T = \{ \tau, w \}$ the tax system, made of the schedule of tax rates $\tau = (\tau_i)_{i=1...n}$, and the potential lump-sum transfer $w$, which could be optimized or not (then $w$ is just a fixed value, perhaps 0). From Proposition 8, define, for a given $\tau^d$:

$$L_i \left( \tau, \theta \right) := \sum_h \left( \lambda - \gamma^h \right) c^i_1 + \sum_h \left[ \gamma^h \tau^s + \left( \lambda - \gamma^h \right) \tau \right] \cdot S^{B,R,h}_i S^{B,R,h}_i = \sum_j S^{r,h}_j M^h_{ji}.$$
contradiction with the optimality of small neighborhood unique and continuous in
de we have $L_i (T, \theta_0) > 0$ whenever $\tau_i \notin T$. This implies that $\tau$ can’t be an optimum; indeed, $\tau_i$ should be raised up to the next quantile in $T$. This proves that any tax $\tau_i$ must have a value in $T$, which has at most $Q$ elements. In short, simple taxes are optimum when $\theta = \theta_0$.

Case $\theta$ close to $\theta_0$. We call $T^*$ the optimal tax system when $\theta = \theta_0$. We reason for $(\tau, \theta)$ in small neighborhood $K$ of $(T^{**}, \theta_0)$. Because $L_i (T^{**}, \theta_0) > 0$, we know that, perhaps by reducing $K$, we have $L_i (\tau, \theta) > 0$ for all $(\tau, \theta) \in K$.

It is clear that for a generic economy, for $\theta$ close enough to $\theta_0$, the optimum tax $T^*(\theta)$ is unique and continuous in $\theta$. So, for $\theta$ close enough to $\theta_0$, $(T^*(\theta), \theta)$ will be in $K$. This implies $L_i (T^*(\theta), \theta) > 0$ for all $i$. Hence, if we had $\tau_i \notin T$, we’d have $\frac{\partial L_i}{\partial \tau_i} (T^*, \theta) = L_i (T^*(\theta), \theta) > 0$, a contradiction with the optimality of $\tau_i$. Hence, we must have $\tau_i \in T$. As in the case $\theta = \theta_0$, this means that all taxes have one of the $Q$ values in $T$: the tax system is simple. \(\square\)

Derivation of relation (42) We take the material from section 9.1, using the budget constraint $B (c, z) = c - qz$. The appearance of a $-q$ requires some care when applying the usual formulas. The good is $C = (c, z)$, and the price $\tilde{q} = (1, q)$, but $B (C, \tilde{q}) = c_1 \tilde{q}_1 - c_2 \tilde{q}_2$. Applying definition (62) gives

$$\tau^b = (1, -q) - \frac{(u_c, u_z)}{v_r}$$

We know that $c_{Qz^*} - qzQ^*_z = 0$ (because $c - qz = w$), so

$$S^{C}_{Qz^*} (c_{Qz^*}, zQ^*_z)' = (q, 1) z'Q^*_z$$

So, from Proposition 24 implies:

$$\frac{v_{Qz^*} (q, w)}{v_w (q, w)} = -\tau^b (q, w) \cdot S^{C}_{Qz^*} (q, w) = -\tau^b (q, w) (q, 1)' zQ^*_z = -\tau^b zQ^*_z$$

$$= -\tau^b \frac{z}{\tilde{q} Q^*_z}$$

as we defined $\tau^b = \tau^b (q, w) \cdot (q, 1) = \frac{u_c + u_z}{v_r}$.

Likewise, $c - qz = w$ implies: $c_q - qz_q - z = 0$ and $c_w - qz_w = 1$, so

$$S^{C}_q (q, w) = C_q - C_w z = (c_q - c_w z, z_q - z_w z)$$

$$= (qz_q + z - (qz_w + 1) z, z_q - z_w z) = (q, 1) (z_q - z_w z) = (q, 1) \frac{zQ^*_c}{q}$$

\(71\)
Proposition 24 implies:

\[ \frac{v_q(q, w)}{v_w(q, w)} = z - \tau^b(q, w) \cdot S^{C}_q(q, w) = z - \tau^b(q, w) \cdot (q, 1) \cdot \frac{c}{q} \]

\[ = z - \frac{z}{q} \tau^{*b} c. \]

\[ \square \]

**Notation for the proofs in the Mirrlees framework**  We find it convenient to define

\[ \overline{D}_{zz} = -\tau^b \frac{c}{q}, \quad \overline{D}_{zz^*} = -\tau^b \frac{c_{Q^*}}{q}. \]

Therefore we have

\[ \frac{v_q^*}{v_w} = z \left(1 + \overline{D}_{zz}\right), \quad \frac{v_{Q^*}}{v_w} = z \overline{D}_{zz^*}. \]

We use the notation

\[ \overline{D}_{zz^*} = \overline{D}_{zz^*} + \overline{D}_{zz} \delta_z(z^*) \]

in the sense of generalized functions.

**Proof of Elasticity relations (43) in the Mirrlees framework**  Now consider the model with misperception. We have \( \frac{v_Q^*}{v_w} = z \overline{D}_{zz^*} = -\tau^b \frac{z c_{Q^*}}{q} \) in the general model. To find the counterpart in the misperception model, we observe that the budget constraint is \( B(c, z) = c - qz \). Hence, Proposition 41 gives: (using \( m_{zz^*} := q^s_{Q^*} (q, Q, r_0) \))

\[ -\tau^b \frac{z c_{Q^*}}{q} = \frac{v_{Q^*}}{v_w} \cdot \frac{v_{Q^*}}{v_w} q^s_{Q^*} = (B^s_{z} - B_z) S^r m_{zz^*} = (-q^s + q) S^r m_{zz^*} \]

The detail of the derivations is as follows:
\[ z_q(q, Q, r_0, r) = \frac{z}{q} \zeta^u = z_q(q, q_0^*, r) + z_q q_q^*(q, Q, r_0) = z_q^* + (S^r + z_r D^s) m_{zz} \] using Proposition 42

\[ = \frac{z}{q} \eta + \left( \frac{z \eta D^s + S}{q} \right) m, \]

\[ z_{Q*} = z_q q_{Q*}^* = (zz_w D^s + S^r) m_{zz*} = \frac{z}{q} \zeta^c \frac{m_{zz*}}{m}, \]

\[ \zeta_{Q*}^c := \frac{q}{z} z_{Q*}^* = c \frac{m_{zz*}}{m}, \]

\[ \zeta^c := \frac{q}{z} z_q - \eta = \zeta^u - \eta = \frac{q}{z} (z_q - z_w z) = \]

\[ = \frac{q}{z} \left[ z - \eta + \left( \frac{z \eta D^s + S}{q} \right) m \right] - \eta = \left( \eta D^s + \frac{q}{z} S \right) m, \]

\[ \zeta^c = \frac{q}{z} (zz_w D^s + S) m = \frac{q}{z} \left( zz_w (q - q^s) \frac{S^r}{z} + S^r \right) m \]

\[ = \frac{q}{z} (z_w (q - q^s) + 1) S^r m. \]

\[ \zeta_{Q*}^c = \frac{\zeta^c m_{zz*}}{m} \]

\[ \zeta^c_{Q*} = \frac{\zeta^c}{D_{zz*}} \]

Note that \( \zeta^c \) has a observable interpretation in terms of compensated price chance, given \( Q \) constant: consider a change \( x \), and \( Z(x) = z^s (q + x, Q, r - z (q, Q, r) x) \). Then, \( \zeta^c = \frac{q}{z} Z'(0) \).

We therefore have

\[ \zeta^c = \left( \eta D^s + \frac{q}{z} S^r \right) m, \]

\[ \zeta^c_{Q*} = \left( \eta D^s + \frac{q}{z} S^r \right) m_{zz*}, \]

where

\[ D^s = (q - q^s) \frac{S^r}{z} = (\tau^s - \tau) \frac{S^r}{z}. \]

Mapping back to the general model, this means that

\[ D_{zz} = D^m m_{zz}, \quad D_{zz*} = D^m m_{zz*}, \]

implying

\[ D_{zz} = (q - q^s) \frac{m S^r}{z} = (\tau^s - \tau) \frac{m S^r}{z}. \]

This leads to

\[ \tau^b = (\tau - \tau^s) \frac{m S^r q}{z} \zeta^c. \]
We have

\[
\begin{align*}
\zeta^c &= \left( \eta \mathcal{D}^p + \frac{q}{z} S^p \right) m, \\
&= \left( \eta \frac{\tau^s - \tau}{z} + \frac{q}{z} S^p \right) m, \\
&= \left( \eta \frac{\tau^s - \tau}{z} + \frac{q}{z} \zeta^c r^b z \frac{1}{q \tau - \tau^s} \right),
\end{align*}
\]

leading to

\[
\tau^b = \frac{\tau - \tau^s}{1 - \eta \frac{\tau - \tau^s}{q}}.
\]

**Misperception model**

Equation (83) and \( q u_c + u_z \) give:

\[
\tau^b = (q, 1) \tau^b = (q, 1) \left[ (1, -q) - \frac{(u_c, u_z)}{v_r} \right]
= - \frac{q u_c + u_z}{v_r} = \frac{u_c - u_z}{v_r u_c u_r}.
\]

**Proof of Lemma 3** Take first the case \( z^* \neq z \). We have

\[
\begin{align*}
z &= z(q(z), Q, r(z)) \\
r(z) &= R(z) - zq(z)
\end{align*}
\]

so

\[
\begin{align*}
\delta r &= -z q'(z) \delta z + \delta K = -z R'' \delta z + \delta K - z \delta q_z \\
\delta z &= z q(q'(z) \delta z + \delta q_z) + \int_0^\infty \zeta_{Q_z} \delta q_z \tau^* dz^* + z_r \delta r \\
&= \frac{z}{q} \zeta^a R'' \delta z + \frac{z}{q} \int_0^\infty \zeta_{Q_z} \delta q_z \tau^* dz^* + \frac{\eta}{q} (-z R'' \delta z + \delta K)
\end{align*}
\]

so

\[
\begin{align*}
\delta z &= \frac{z}{q} \int_0^\infty \zeta_{Q_z} \delta q_z \tau^* dz^* + \eta \delta K \\
&= \frac{z}{q} \int_0^\infty \zeta_{Q_z} \delta q_z \tau^* dz^* + \eta \delta K
\end{align*}
\]
For welfare $v^* (q, Q, r_0, r)$, we have:

$$
\frac{\delta v}{v_w} = \frac{v^*_Q}{v_w} \delta q_z + \int_0^\infty \frac{v^*_Q}{v_w} \delta q_z^* dz^* + \delta r
$$

$$
= (z + z \tau \xi (z)\delta z + \delta q_z) + z \int_0^\infty \tau \xi (z)\delta q_z^* dz^* = z R'' \delta z + \delta K - z \delta q_z
$$

$$
\frac{\delta v}{v_w} = z \tau \xi (z)\delta z + z \left( \int_0^\infty \tau \xi (z)\delta q_z^* dz^* + \int_0^\infty \tau \xi (z)\delta q_z dz \right) + \delta K
$$

$$
= \delta K - z \tau \xi (z)\delta z + \int_0^\infty \tau \xi (z)\delta q_z^* dz^*
$$

\(\square\)

**Proof of Lemma 4**  Observe that

$$
\frac{h^* (z)}{q} = \frac{h (z)}{q - \xi z R'' (z)}
$$

so that $q - \xi z R'' (z) = \frac{qh}{h^*}$ and

$$
z R'' = q \frac{h^* - h}{\xi h^*}.
$$

(84)
We have

\[ \delta L = T'(z) \delta z + g(z) \frac{\delta v}{v_w} - \delta K \]

\[ = T'(z) \delta z + g(z) \left( \overline{D}_{zz} z R'' \delta z + z \int_0^\infty \overline{D}_{zz} \delta q_z \, dz^* + \delta K \right) - \delta K \text{ by Lemma 3} \]

\[ = \left( T'(z) + g(z) \overline{D}_{zz} \frac{h^* - h}{h} \right) \delta z + g(z) \left( z \int_0^\infty \overline{D}_{zz} \delta q_z \, dz^* + \delta K \right) - \delta K \text{ by (84)} \]

\[ = \left( T'(z) + g(z) \overline{D}_{zz} \frac{h^* - h}{h} \right) \delta z + g(z) \left( z \int_0^\infty \overline{D}_{zz} \delta q_z \, dz^* + \delta K \right) - \delta K \text{ by Lemma 3} \]

\[ = \left( T'(z) + g(z) \overline{D}_{zz} \frac{h^* - h}{h} \right) \delta z + g(z) \left( z \int_0^\infty \overline{D}_{zz} \delta q_z \, dz^* + \delta K \right) - \delta K \]

Note that we can also express:

\[ \gamma(z) = g(z) + \frac{T'(z)}{q} \eta h^*(z) + g(z) \overline{D}_{zz} \eta \frac{h^*(z) - h(z)}{\zeta h(z)} \]

\[ = g(z) + \frac{T'(z)}{q} \eta h^*(z) - g(z) \frac{\tau_b h^*(z) - h(z)}{q} \]

\[ = g(z) + \frac{T'(z) h^*(z)}{q} - g(z) \frac{\tau_b h^*(z) - h(z)}{q} \]

\[ = g(z) + \eta g(z) \frac{\tau_b h(z)}{q} + \frac{T'(z) h^*(z)}{q} \frac{\tau_b h^*(z) - h(z)}{h(z)} \]

\[ = g(z) + \eta \frac{\tau_b h(z)}{q} + \frac{T'(z) h^*(z)}{q} \frac{\tau_b h^*(z) - h(z)}{h(z)} \]

\[ = g(z) \left( \frac{T'(z) h^*(z)}{q} + \frac{\tau_b h^*(z) - h(z)}{h(z)} \right) \]

\[ = \left[ g(z) h(z) + \frac{T'(z)}{q} \eta h^*(z) + g(z) \overline{D}_{zz} \eta \frac{z R''}{q} h^*(z) \right] / h(z) \]

Proof of Proposition 17  We use the notations
\[
\overline{F}(z_\ast) = \int_0^\infty \xi_{Q_\ast} \frac{T'(z) - \overline{T}'(z)}{q(z)} \, zh^*(z) \, dz \\
= \int_0^\infty \left[ g(z) \overline{D}_{zz_*} + \xi_{Q_\ast} \frac{T'(z)}{q} \right] \, zh^*(z) \, dz = \overline{F}(z_\ast) + J(z_\ast)
\]

\[
\overline{F}(z^*) := \int_0^\infty \left[ g(z) \overline{D}_{zz_*} + \xi_{Q_\ast} \frac{T'(z)}{q} \right] \, zh^*(z) \, dz \\
J(z^*) := \xi_{z^*} \frac{T'(z^*)}{q} \, h^*(z^*)
\]

\(\overline{F}(z^*), \overline{F}(z_\ast)\) and \(J(z^*)\) are generalized functions of \(z^*\).

We consider a change \(\delta q_{z^*}\) at \(z = z^*\). This leads to a lump-sum change \(\delta K = 1_{z > z^*} \delta q_{z^*}\). Hence, Lemma 4 gives the changes:

\[
\delta L(z) = (\gamma(z) - 1) 1_{z > z^*} \delta q_{z^*} + \left[ g(z) \int_0^\infty \overline{D}_{zz_*} \delta q_{z^*} \, dz^* + \frac{T'(z)}{q} \int_0^\infty \xi_{Q_\ast} \delta q_{z^*} \, dz^* \right] \frac{h^*}{h} 
\]

the \(-1_{z > z^*} dP\) term comes from the lost revenue to the government. The total change is

\[
\frac{\partial L}{\partial \tau_{z^*}} = -\overline{F}(z_\ast) + \int_{z^*}^\infty (1 - \gamma(z)) h(z) \, dz 
\]

Taking out the Diracs from \(\overline{D}_{zz_*}, \xi_{Q_\ast}\), we get

\[
\frac{\partial L}{\partial \tau_{z^*}} = -\int_0^\infty \left[ g(z) \overline{D}_{zz_*} + \xi_{Q_\ast} \frac{T'(z)}{q} \right] \, zh^*(z) \, dz + \int_0^\infty \left( 1 - g(z) - \frac{T'(z)}{q} \frac{h^*}{h} - g(z) \overline{D}_{zz_*} \frac{h^*}{\xi} h \right) \, h(z) \, dz \\
= -F(z_\ast) - J(z_\ast) - \xi_{Q_\ast} \overline{D}_{zz_*} h^*(z^*) + \int_{z^*}^\infty \left( (1 - g(z) - \lambda \delta h - g(z) \overline{D}_{zz_*} \lambda (h^* - h) \right) \, dz \\
= -\overline{F}(z_\ast) - J(z_\ast) + \int_{z^*}^\infty (a(z) - \lambda \delta h) \, dz
\]
Proof of Proposition 22  Case of an inattentive consumer. Call \( q_a = p_a + \tau_a \). Equilibrium requires \( q_a = q_b = 1 \). Competitive pricing in good 1 requires that firms choose inputs according to: \[
\max_{l_{ia},l_{ib}} p_i \left( \frac{l_{ia}}{\alpha_i} \right)^{\frac{\alpha_i}{1-\alpha_i}} - (1 + \tau_{ia}) l_{ia} - l_{ib} \text{ with } \tau_{0a} = 0.
\]
Hence, the equilibrium price is \( p_i = (1 + \tau_{ia})^{\alpha_i} \), and input use features \( (1 + \tau_{ia}) l_{ia} = \alpha_i p_i y_i \), so \( l_{ia} = \alpha_i (1 + \tau_{ia})^{\alpha_i-1} y_i \) and \( l_{ib} = (1 - \alpha_i)(1 + \tau_{ia})^{\alpha_i} y_i \).

The planning problem is \( \max_{\tau_{ia}} L(p_1) \) with \( p_1 = (1 + \tau_{ia})^{\alpha_1} \), so that:

\[
L(p_1) = c_0 + U^b(c_1(p_1)) - \xi_s c_1(p_1) - \sum_{i=0}^{1} (l_{ia}(p_1) + l_{ib}(p_1))
\]

\[
= U^b(c_1(p_1)) - \left( \xi_s + \alpha_1 (1 + \tau_{ia})^{\alpha_1-1} + (1 - \alpha_1) (1 + \tau_{ia})^{\alpha_1} \right) c_1(p_1) \text{ as } c_0 - (l_{0a}(p_1) + l_{0b}(p_1))
\]

\[
= U^b(c_1(p_1)) - \left( \xi_s + \alpha_1 p_1^{1 - \frac{1}{\alpha_1}} + (1 - \alpha_1) p_1 \right) c_1(p_1)
\]

with \( c_1(p_1) = U^{r-1}(p_1) \).

Hence, as \( U^{br}(c_1(p_1)) = p_1 \),

\[
L_{p_1} = c'_1(p_1) \left[ p_1 - \left( \xi_s + \alpha_1 p_1^{1 - \frac{1}{\alpha_1}} + (1 - \alpha_1) p_1 \right) \right] - \left( (\alpha_1 - 1) p_1^{\frac{1}{\alpha_1}} + (1 - \alpha_1) \right) c_1(p_1)
\]

When there is production efficiency, \( p_1 = 1 \) and,

\[
L_{p_1|\tau_{ia}=0} = c'_1(p_1) \left[ U^{br}(c_1(p_1)) - (\xi_s + 1) \right]
\]

\[
= -\xi_s c'_1(p_1) > 0
\]

Hence, production efficiency is not an optimum. Starting from it, it is optimal to increase the tax \( \tau_{ia} \) to discourage the production of good 1, increase its price, and discourage its consumption.

Case of an attentive consumer. It is enough to do a Pigouvian tax \( \tau^c = \xi_s \), and restore production efficiency \( (\tau_{ia} = 1) \). Then, we achieve the first best. \( \square \)