Inattention and the Taxation Bias*

Jeremy Boccanfuso[†] Antoine Ferey[‡]

February 15, 2019

Abstract

This paper studies how information frictions in tax perceptions interact with the design of tax policy. It develops a positive theory of tax policy in which taxpayers endogenously decide how much attention to devote to observing taxes and the government optimally sets tax policy taking taxpayers' attention choices into account. In equilibrium, we show that the government faces a commitment problem and implements inefficiently high tax rates. This *taxation bias* directly depends on agents' attention and sheds a new light on the potential welfare implications of tax misperceptions. Most notably, we find that the widely documented underestimation of tax rates may not be welfare-improving if taxpayers are not sufficiently attentive to tax policy.

Keywords: Optimal taxation; inattention. JEL classification code: H21; D03.

*We thank Pierre Boyer, Philippe Choné, Carlos Da Costa, Xavier Gabaix, Louis Kaplow, Etienne Lehmann, Benjamin Lockwood, Jean-Baptiste Michau, Xavier Ragot, Anasuya Raj, Daniel Reck, Alessandro Riboni, Emmanuel Saez, Stefanie Stantcheva, Dmitry Taubinsky, Nicolas Werquin and audiences of the AFSE Conference 2018, Berkeley, CREST, CRED Taxation Group, LAGV Conference 2018, NTA Conference 2018 and PET Conference 2017 for very helpful discussions and feedback. Any remaining errors are our own.

[†]PSE – EHESS, 48 Boulevard Jourdan, 75014 Paris, France. E-mail: jeremy.boccanfuso@psemail.eu

[‡]École Polytechnique – CREST, Route de Saclay, 91128 Palaiseau, France. E-mail: antoine.ferey@polytechnique.edu

1 Introduction

Tax systems are notoriously complex and hard to understand. As a result, a growing body of evidence documents information frictions in tax perceptions such as misperceptions about, or inattention to, tax rates (e.g. Chetty et al. (2009), Rees-Jones and Taubinsky (2016)). More specifically, the bulk of the evidence shows taxpayers tend to underestimate tax rates and under-react to variations in the latter. This suggests that, by reducing tax distortions and their associated efficiency costs, these information frictions may be welfare improving.

In this paper, we demonstrate that information frictions create an incentive for the government to increase tax rates beyond efficient levels. This *taxation bias* directly depends on agents' attention and thus on the magnitude of information frictions. By highlighting the existence of an additional source of inefficiency, our analysis sheds a new light on the global welfare implications of these information frictions and challenges the idea that potential welfare gains may actually be attained.

To formalize these arguments, we consider the problem of a taxpayer who chooses how much to earn and consume in the presence of an income tax. Because of information frictions, she cannot freely observe the tax policy implemented by the government but has an exogenous prior, or default, about it. We remain voluntarily agnostic about the origins of this prior and consider the possibility that it may be biased. It thus allows to consider various sources of tax misperceptions including, but not limited to, salience (e.g. Chetty et al. (2009); Gabaix (2014)), linearizing heuristics (Liebman and Zeckhauser (2004); Rees-Jones and Taubinsky (2016)) or experienced taxes. To improve her perception of the tax policy a taxpayer may in addition acquire costly information in the form of a Gaussian signal with endogenous precision which reflects taxpayer's attention. Using Bayes' rule, she then forms an estimate of the actual tax by combining her prior and the acquired information. As a result, taxpayers' perceptions are anchored on the prior and adjustments to unanticipated tax changes are proportional to their choice of attention.

The government is a social welfare maximizer who sets a linear tax policy[1] taking taxpayers choices of attention, earnings and consumption into account. Importantly, agents' inattention to taxes results in a partial anchoring of their perceptions which dampens their responsiveness to tax reforms. This *elasticity effect* creates an incentive for the government to increase taxes in order to collect higher tax revenues at low efficiency costs. Intuitively, information frictions allow for policy deviations in which the government fools the agents.[2]

[1] We consider a linear tax policy because it substantially simplifies the analytical treatment of agents' perceptions model. An extension to non-linear taxes is to be added in a future version of the paper.

[2] Note that the government is here fundamentally paternalistic and tries to fool the agents for their

The key novelty of the paper is thus to develop a positive theory of tax policy to study how information frictions in tax perceptions interact with the design of tax policy. This approach seems particularly relevant to analyze a situation in which the government has a fundamental incentive to exploit taxpayers biases or, in other words, to fool them. Given their prior and the uncertainty[3] about tax policy, agents decide on their attention strategies. Given taxpayers' prior, the government can infer agents' attention strategies and take them into account when setting tax policy. In equilibrium, attention strategies are optimal, tax policy is optimal and priors are adjusted up to an arbitrary perception bias.

When setting the tax policy, the government does not internalize the equilibrium consequences of its choice. This source of inefficiency generates a taxation bias. Intuitively, the government anticipates tax increases to have a low efficiency costs thanks to the elasticity effect. This elasticity effect however vanishes in equilibrium. The economy therefore ends up in a situation where the tax rate is higher than optimal. This taxation bias fundamentally arises because the government cannot credibly commit to a lower tax policy as it would always have an incentive to deviate ex post. Consequently, information frictions on the taxpayers's side lead to a commitment problem similar to that encountered by central bankers (e.g. Barro and Gordon (1983)).

To demonstrate these results, we proceed in two steps. First, we analytically characterize the equilibrium policy. Further using a tractable Gaussian learning model, we provide a simple sufficient statistics formula. Second, we characterize the normative policy that would be chosen by a benevolent social planner who accounts for the equilibrium effects of its choice and derive an analogously simple formula. Using these simple formulas, we provide a closed-form expression for the taxation bias defined as the difference between the positive and the normative tax rate. We find that it directly depends on agents' attention, their responsiveness to taxes and the government's redistributive tastes.

The overall welfare impact of information rigidities may then be decomposed between potential welfare gains from the normative equilibrium and the welfare costs induced by the taxation bias. Using the tractable Gaussian learning model, information rigidities are shown to be unambiguously detrimental to welfare in the absence of perception bias, even though the potential welfare gains might be positive. Surprisingly, when agents' prior is downward biased in equilibrium – a situation in line with the empirical evidence

own good, or more precisely for society's own good.

[3]Formally, we introduce implementation shocks reflecting the uncertainty in the political process leading to the implementation of tax policy which may prevent the government to implement its preferred tax policy. This should be seen as a transparent way to introduce fundamental uncertainty in tax policy. A similar approach is taken in Matějka and Tabellini (2017). We provide a discussion of this point in Section 3.2.2.

on the underestimation of tax rates – the welfare impact of information rigidities could go either way. Indeed, the potential welfare gains are an increasing and concave function of information costs, while the welfare costs from the taxation bias are increasing and convex. As a result, small misperceptions compatible with high levels of attention in equilibrium are likely to be welfare improving. However, large misperceptions associated to low levels of attention will typically be welfare decreasing.

Related literature This paper is deeply connected to the growing literature in behavioral public economics (see the reviews of Chetty (2015) and Bernheim and Taubinsky (2018)) and first and foremost to the normative literature on optimal taxation with behavioral agents (e.g. Farhi and Gabaix (2015), Gerritsen (2016), Lockwood (2016)). These papers usually optimize tax policy taking the impact of behavioral biases on labor supply choices into account. Formally, they rely, as we do, on labor supply models in which misperceptions affect earnings choices and where consumption adjusts such that agents' budget constraints are always satisfied.[4] They then show how a social planner can optimally exploit behavioral biases to increase social welfare, which corresponds to our normative benchmark. Importantly, the presence of behavioral biases potentially allows to improve upon the standard second-best allocation in which agents are rational and have full information (Mirrlees (1971) and Saez (2001)), suggesting that these behavioral biases could be welfare improving.

However, this normative literature is mostly silent on a key aspect of the problem: "A difficulty confronting all behavioral policy approaches is a form of Lucas critique: how do the underlying biases change with policy?" [Farhi and Gabaix (2015), page 13]. Our first contribution to this literature is to construct a general and flexible framework that explicitly accounts for perceptions' adjustments through two channels: adjustment in the priors and adjustment in attention strategies. Our second contribution is to develop a positive model of tax policy which takes the form of a game in which the government acts as an economic agent taking decentralized decisions and perceptions are adjusted in equilibrium. This game theoretic approach seems particularly relevant to model a situation in which the government fundamentally tries to exploit taxpayers biases, or in other words, to fool the agents. As a result, the government fails to internalize part of the adjustment in its choice which creates a new source of inefficiency. Even behavioral biases that are *prima facie* welfare-improving may actually end up being welfare-decreasing in equilibrium: third-best allocations featuring this new inefficiency do not necessarily improve upon standard second-best allocations. Fundamentally, this is a commitment problem. Such behavioral biases would be welfare-improving if the government could

[4]See Reck (2016) for a discussion of budget adjustment rules in misperception models

commit to the normative policy but this is not credible given its incentive to deviate. Because of the parallel with the model of Barro and Gordon (1983), we call this inefficiency the *taxation bias*.

Second, our work is related to the recent body of empirical work documenting substantial inattention to, or misperceptions about, taxes. We contribute to this literature by improving the understanding of the welfare implications of these misperceptions and highlighting their potentially important equilibrium effects. In a seminal paper, Chetty et al. (2009) show that in the context of sales tax consumers exhibit a salience bias: they are inattentive to taxes that are not included in posted prices and thereby less responsive to them. Using an online shopping experiment, Taubinsky and Rees-Jones (2017) confirm these findings and show in addition that inattention is heterogeneous across consumers and responds to tax incentives. In their setting, tripling the tax rate nearly doubles consumers' attention. Contrasting the impact of salient and non-salient changes in the subsidization of child and dependent care in the US, Miller et al. (2015) show that these salience effects are largely portable to income tax settings. More specific to the context of income taxes, Liebman and Zeckhauser (2004) argue that taxpayers use a linearizing heuristic called ironing which consists in approximating their marginal tax rates by their average tax rates. In a lab experiment Rees-Jones and Taubinsky (2016) elicit the income tax perceptions of American taxpayers and document the widespread use of this heuristic. Overall, these findings motivate our flexible perceptions formation model. It relies on the combination of an exogenous prior, or default, accommodating a wide range of biases, heuristics or beliefs and an endogenous information acquisition process capturing agents' responses to variations in tax incentives. Endogenous information acquisition here generates a positive correlation between taxpayers' attention and income, as observed in the data by Taubinsky and Rees-Jones (2017). Moreover, the presence of noisy information generates dispersion and heterogeneity in perceptions.[5]

Finally, the paper builds on recent contributions in inattention. We use a noisy information model with Gaussian signals which precisions are endogenous. Further, we consider a Bayesian learning to form an estimate of the tax rate. Imposing this structure avoids technical questions – such as finding the expected utility maximizing filtration of the signals or the signal distribution given the information cost – that would otherwise be central in, for example, the rational inattention literature (Sims (2003)). This simplification allows the use of a tractable and transparent learning model that has been widely used in economics.[6] Thus far, the behavioral public finance literature has mainly consid-

[5]Adding more dispersion in the model could be done through individual-specific priors, but it would ultimately require individual-specific sources of uncertainty to formally close the model i.e. a rationale explaining why these individual-specific priors differ.

[6]See for example the reviews in Gabaix (2019) section 6 and Mackowiak et al. (2018) .

ered deterministic forms of inattention – e.g. salience (Chetty et al. (2009)), linearized heuristics (Liebman and Zeckhauser (2004)), experienced utility and misperception model (Farhi and Gabaix (2015)). We depart from this approach and consider an inattention model that generates stochasticity in actions, which appears to be coherent with the observed dispersion and heterogeneity in perceptions. We nevertheless allow perceptions to be biased in equilibrium through an arbitrary bias in the prior. Doing so is a way to build a bridge between these previous studies and our own. Finally, we follow Sims (2003) and measure the cost of information from the expected entropy reduction from observing a signal. Fundamentally, the presence of the taxation bias does not follow from a specific model of attention. As is shown in the subsequent section which discusses a toy example, it is robust to an alternative attention modeling leading to deterministic actions.

The paper is organized as follows. It first presents the taxation bias in a simple toy model. Section 3 then focuses on the behavior of taxpayers in the economy. Sections 4 and 5 respectively characterize the positive and normative equilibria along with the taxation bias. Finally, the last section analyzes welfare implications.

2 Toy model

This section discusses a voluntarily simplistic toy model to motivate the main result of the paper. It aims at exposing the underlying economic mechanisms leading to the taxation bias in a straightforward fashion. The rest of the paper formalizes these mechanisms in a richer and fully micro-founded environment.

Consider a canonical labor income taxation model where the government sets a linear tax rate τ to maximize tax revenue.[7] We depart from standard assumptions only in the introduction of information frictions making taxpayers unable to freely observe the tax rate. Consequently, they must form an estimate of the latter when deciding how much to work. For ease of exposition, let the perceived tax rate $\tilde{\tau}$ be a convex combination of a common prior $\hat{\tau}$ and the actual tax rate τ. Namely, $\tilde{\tau} = \xi\tau + (1-\xi)\hat{\tau}$ with $\xi \in [0,1]$ a measure of attention to the actual tax rate. Further define an aggregate earnings function $Y(\tilde{\tau})$ which depends on this perceived tax rate and $e = \frac{1-\tilde{\tau}}{Y}\frac{\partial Y}{\partial(1-\tilde{\tau})} > 0$ the elasticity of aggregate earnings with respect to the perceived net-of-tax rate. In the canonical model with full information, the tax revenue function is $\tau Y(\tau)$. It has an inverted U-shape and is nil when τ is equal to either 0 or 100%. In our set-up with information frictions, the tax revenue function becomes $\tau Y(\tilde{\tau})$ and is still concave. Appendix A.1 explicitly solves this

[7] See for example Piketty and Saez (2013), section 4.1, pages 410-411.

taxation problem. In the following, we discuss and interpret the implications of imperfect information in an intuitive way.

How do information frictions interact with the design of tax policy? When setting the tax rate, the government simply aims to reach the top of the Laffer curve. A marginal increase in $d\tau$ lowers aggregate earnings by $\frac{\partial Y}{\partial \hat{\tau}}\xi d\tau$ – because taxpayers observe only a fraction ξ of the tax increase $d\tau$ – and increases tax revenue by $Y d\tau$. Hence, a direct consequence of information frictions is to reduce the responsiveness of aggregate earnings to the tax rate. In comparison to a situation with perfect information, the top of the Laffer curve shifts to the right and the government sets a higher tax rate. More specifically, the revenue maximizing tax rate decreases with the elasticity of aggregate earnings and taxpayers' prior but increases with inattention:

$$\tau = \begin{cases} 1 & \text{if} \quad \hat{\tau} \leq 1 - \frac{\xi}{1-\xi} e \\ \frac{1}{1+\xi e} & \text{otherwise} \end{cases} \tag{1}$$

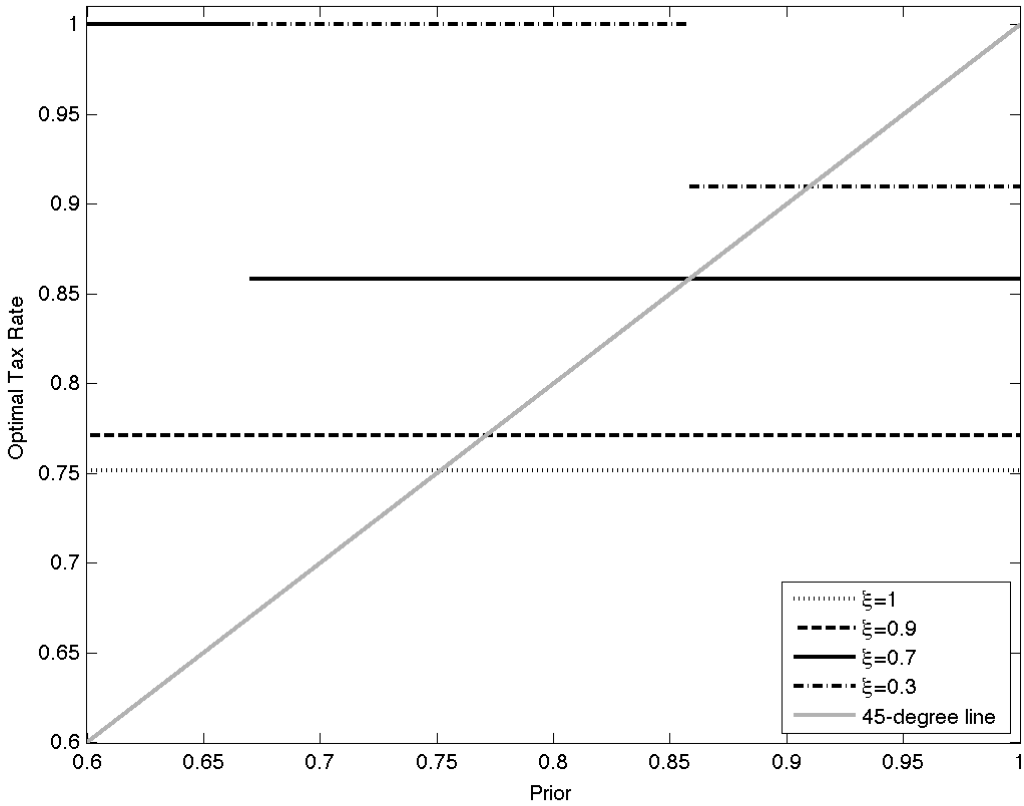
To better understand the government incentive to deviate from the optimal policy with perfect information, assume that taxpayers initially anchor their expectations at $\hat{\tau} = \tau^n$ where $\tau^n \equiv \frac{1}{1+e}$ is the revenue maximizing tax rate under perfect information ($\xi = 1$). Then, the government has an incentive to fool taxpayers as they only internalize a fraction ξ of the tax increase above $\hat{\tau}$. As a result, setting $\tau = \tau^n$ is no longer optimal and the government targets a higher tax rate. Ultimately, the government raises τ up to the point where the marginal gains from fooling taxpayers have been exhausted.

Figure 1 reports the optimal policy when $e = 0.33$ (Chetty, 2012). Small information rigidities may generate substantial deviation in tax policy. When 90% of taxpayers' perceived rate is driven by the true rate and inattention is minimal, the optimal tax rate is more than 2 percentage points higher than what it would be in the absence of information frictions.

However, taxpayers cannot be systematically fooled and their prior must adjust once they understand the government's incentive. In equilibrium, taxpayers' prior must be consistent with the policy chosen by the government. Among possible equilibria, the rational expectation equilibrium is of special interest here. It implies that expectations are fully adjusted, that is $\hat{\tau} = \tau$ and as demonstrated in Appendix A.1, it is an interior solution to the problem. Graphically, the rational expectation equilibrium corresponds to the point where the 45-degree line crosses the government policy function in figure 1.

In the presence of information frictions on tax perceptions, the government is unable to reach the top of the Laffer curve. The tax rate is too high and taxpayers' labor supply too low: the rational expectation equilibrium is strictly dominated by the equilibrium in which the government credibly commits to τ^n. However, the government may not credibly anchor expectations at τ^n since it will always have an incentive to deviate ex

Figure 1: Optimal Taxation in The Toy Model



NOTE: Optimal policy as a function of the prior $\hat{\tau}$ for different values of the misperception parameter ξ. The elasticity of aggregate earnings with respect to the perceived net-of-tax rate is set to 0.33.

post. We call this phenomenon the *taxation bias* in analogy to the *inflation bias* (Barro and Gordon (1983)).

3 Individuals' behavior

This section describes the behavior of taxpayers in the economy. Because of information frictions, taxpayers may not freely observe the tax rate implemented by the government. They rely on a Bayesian learning model with costly information acquisition to form expectations about the tax schedule in order to decide how much to earn and consume.

3.1 Primitives and assumptions

We consider a population of agents with heterogeneous productivities w which are private information and distributed from a well-defined probability distribution function $f_w(w)$. Each taxpayer has a utility function $U(c, y; w)$ where c is consumption and y earnings.

$U(.)$ is continuously differentiable, increasing in consumption ($U_c > 0$), decreasing in effort ($U_y < 0$ and $U_w > 0$) and such that the Spence-Mirrlees condition holds (MRS_{yc} decreases with skill w).

Agents choose their consumption c and earnings y subject to an income tax $T(y)$. Because of information frictions, taxpayers are however unable to freely observe $T(y)$. They instead use an individual specific perceived income tax, denoted $\tilde{T}(y)$. We jointly model the choice of allocations (c, y) and tax perceptions $\tilde{T}(y)$ under the following assumptions:

Assumption 3.1 (linear representation). *Individuals use a linear representation of the tax schedule* $\tilde{T}(y) = \tilde{\tau}y - \tilde{R}$

Assumption 3.2 (linear tax schedule). *The government sets a linear tax rate* τ_0 *and redistributes through a lump-sum demogrant* R_0 *such that* $T(y) = \tau_0 y - R_0$

Assumption 3.3 (quasi-linear utility). *The utility function is quasi-linear* $U(c, y; w) = c - v(y; w)$

In order to make their consumption and earnings choices, individuals must rely on a perception of the overall tax function, namely an estimate of the tax liability owed at each earning level. Assumption 3.1 imposes that taxpayers use a linear representation of the tax schedule. Consequently, they only need to form estimates of the marginal tax rate $\tilde{\tau}$ and the intercept $\tilde{R}$. This can be interpreted as a local first-order approximation of a potentially non-linear tax schedule. Beyond the fact that a linear approximation is usually a good first-order approximation of existing tax schedules (Piketty and Saez, 2013), recent empirical evidence suggests that in practice taxpayers tend to use linear representations of tax schedules (Rees-Jones and Taubinsky, 2016). While Assumption 3.1 is central in reducing the dimensionality of the learning problem to make it tractable, Assumptions 3.2-3.3 are adopted because they allow for a great simplification of the problem but may be relaxed to consider income effects and non-linear tax schedules.[8] In the following, we denote by (τ_0, R_0) the actual tax parameters and by (τ, R) the associated random variables from the point of view of the agent.

3.2 Individual problem

Individuals jointly choose an allocation (c, y) and how much information to collect about the tax schedule. We decompose and solve this problem in two steps. First, we characterize the allocation choice given a perceived tax schedule $\tilde{T}(y)$. Then, we focus on information acquisition taking into account how misperceptions affect allocations.

[8] These extensions are to be added in a future version of the paper.

3.2.1 Allocation choice

Agents choose consumption c and earnings y subject to their perceived budget constraint which depends on their perceptions of the tax schedule. Under Assumption 3.1, this problem writes

$$\begin{aligned} \max_{c,y} \quad & \int_\tau U(c, y; w)\ \tilde{q}(\tau) d\tau \\ \text{s.t.} \quad & c \leq R + (1-\tau)y \end{aligned} \tag{2}$$

where $\tilde{q}(\tau)$ is the perceived distribution of the marginal tax rate τ. From the quasi-linear assumption, there are no income effects and the first-order condition that determines earnings is

$$\frac{\partial v(y; w)}{\partial y} = 1 - \tilde{\tau} \tag{3}$$

with $\tilde{\tau} \equiv E_{\tilde{q}(\tau)}[\tau]$ the average perceived marginal tax rate. Consequently, the average perceived marginal tax rate $\tilde{\tau}$ is a sufficient statistics for labor supply and uniquely pins down earnings $y^\star(\tilde{\tau}; w)$. Hence, a direct implication of Assumption 3.3 is that tax liability, and in particular the perceived value of the demogrant $\tilde{R}$, is irrelevant for labor supply and only matters to determine agents' consumption levels.

Assumption 3.4 (slack budget)**.** *Consumption adjusts such that agents exhausts their true budget i.e.* $\tilde{R} = R_0 - (\tau_0 - \tilde{\tau})y^\star(\tilde{\tau}; w)$

We assume consumption adjusts such that the true budget constraint holds ex post. Although somewhat debatable, this assumption is used throughout the behavioral tax literature as emphasized in **?** who discusses different budget adjustment rules in misperception models. Assumption 3.3 – together with Assumption 3.4 – thus implies that the only parameter of interest for agents' allocation choice is the marginal tax rate τ. It thus significantly simplifies the perceptions formation problem by restricting it to a single parameter.

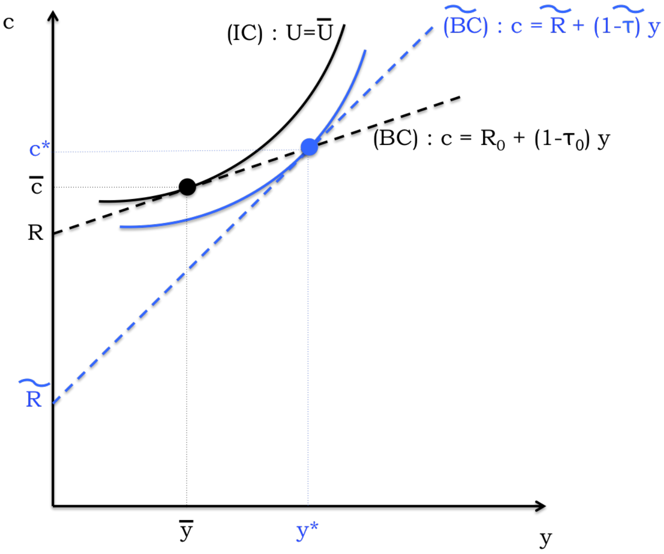
Agent's indirect utility is then

$$V(\tilde{\tau}, \tau_0, R_0; w) = R_0 + (1-\tau_0)y^\star(\tilde{\tau}; w) - v\big(y^\star(\tilde{\tau}; w); w\big) \tag{4}$$

Figure 2 summarizes the allocation choice in a y-c diagram. Perceptions of the tax schedule determine earnings (tangency condition with perceived budget line) while consumption adjusts to the true budget constraint (intersection with true budget line).

A natural observation from Figure 2 is that misperceptions induce *utility misoptimization costs*: the utility level $\bar{U}$ associated with the optimal choice (black dot) is higher than the utility level $U^\star$ associated with the choice under misperceptions (blue dot). However, when agents underestimate tax rates ($\tilde{\tau} < \tau_0$) misperceptions also induce *efficiency*

Figure 2: Allocation choice in a y-c diagram



NOTE: The blue (resp. black) dot represents the allocation choice of an individual who misperceives (resp. correctly perceives) the tax schedule.

gains: labor supply $y^\star$ chosen with misperceptions is greater than the labor supply $\bar{y}$ from the optimal allocation. It appears that efficiency gains dominate utility misoptimization costs. Graphically, the difference in utility which is here money metric (Assumption 3.3) is smaller than the difference in earnings. Hence, if efficiency gains can be redistributed, the underestimation of taxes increases social welfare.

Formally, if we define welfare as the sum of individual utility and tax revenue $W = V(\tilde{\tau}, \tau_0, R_0, w) + \tau_0 y^\star(\tilde{\tau}, w)$, we have that

$$\frac{dW}{d\tilde{\tau}} = (\tilde{\tau} - \tau_0)\frac{dy^\star}{d\tilde{\tau}} + \tau_0 \frac{dy^\star}{d\tilde{\tau}} = \tilde{\tau}\frac{dy^\star}{d\tilde{\tau}} \leq 0 \tag{5}$$

meaning that social welfare increases when individuals' perceived tax rate $\tilde{\tau}$ decreases, regardless of the true tax rate τ_0. Intuitively, distortions on labor supply induced by taxes gradually disappear as the perceived tax rate decreases. It follows that under Assumptions 3.1 - 3.4, tax misperceptions are welfare improving if and only if individuals underestimate marginal tax rates. Given the recent empirical contributions suggesting it may indeed be the case, this line of reasoning suggests that tax misperceptions may be welfare-improving. We come back to this claim when we discuss welfare implications in Section 6.

3.2.2 Perceptions formation

We apprehend tax perceptions as resulting from a combination between an exogenous and free prior (also referred to as a default or an anchor) and an endogenous and costly information acquisition process.[9] The prior accounts for sources of structural and subjective uncertainty.[10] In the subsequent subsections, we voluntarily remain agnostic about these sources of uncertainty for three reasons. First, while the empirical literature clearly indicates that taxpayers misperceive tax rates, there is yet no consensus on the exact rationale (or rationales) behind these misperceptions. Second, the assumed ex ante uncertainty essentially represents a motive for taxpayers to learn in our setup. Therefore, the main results of the paper will hold for a wide variety of priors. Finally, the introduction of exogenous priors allows to consider diverse forms of beliefs and equilibrium concepts, ranging from situations where priors are correct on average to situations where they are biased due to cognitive biases (Chetty et al. (2009), Rees-Jones and Taubinsky (2016)) or motivated beliefs (Benabou and Tirole, 2006).

Let $\hat{q}(\tau)$ be the prior distribution and $\hat{\tau} \equiv E_{\hat{q}}[\tau]$ the expected tax rate derived from the prior. An agent observes the realization of an unbiased Gaussian signal s with variance σ^2 about the actual tax rate τ_0. Based on this signal, she forms a posterior $\tilde{q}(\tau|s;\sigma)$ using Bayes' theorem

$$\tilde{q}(\tau|s;\sigma) \quad \propto \quad \phi(s;\tau,\sigma^2)\hat{q}(\tau) \tag{6}$$

where $\phi(s;\tau,\sigma^2)$ is the gaussian pdf with mean τ and variance σ^2. Building on the rational inattention literature (Sims (2003)), the information content transmitted through the signal is measured from the entropy reduction between the prior and the posterior

$$\mathcal{I}(\sigma) \equiv H(\hat{q}(\tau)) - E_{p(s)}\Big[H(\tilde{q}(\tau|s;\sigma))\Big] \tag{7}$$

where $H(q(\tau)) \equiv -\int q(\tau)\log_2(q(\tau))d\tau$ is the differential entropy (in bits) of the probability distribution $q(\tau)$ and $E_{p(s)}[.]$ the expectation taken over the marginal distribution $p(s) \equiv \int \phi(s;\tau,\sigma)\hat{q}(\tau)d\tau$. The utility cost to acquire and process a unit (bit) of information is constant and denoted κ. The attention strategy of a taxpayer with productivity w thus results from an arbitrage between improved private decisions thanks to more

[9] In a recent survey, Gabaix (2019) argues this is a unifying framework to model various behavioral biases and attention theories. As illustrated by the toy model in section 2 which relies on a simple and transparent expectation formation process, the main result of the paper is robust to various models of perceptions formations. Nevertheless, for the sake of concreteness and clarity, we henceforth consider a Bayesian learning model with a choice of information (in Gabaix (2019) terminology). We choose this model for its wide use in economics, its well-understood micro-foundations and the fact that, as we show, it generates predictions in line with the empirical evidence.

[10] For example, related to policy primitives (e.g. hidden tastes for redistribution), economic fundamentals (e.g. shocks to the government expenditure requirements), institutions (e.g. inability to implementation a chosen policy), heuristic decision rules (e.g. use a linear approximation of non-linear taxes), model uncertainty, etc.

accurate information and the cost to acquire this information. More specifically, she chooses the signal precision – or equivalently the variance $\sigma^\star(\hat{q}(\tau), \kappa, w)$ – to maximize her expected indirect utility

$$\max_{\sigma} \iint V\Big(\tilde{\tau}(s,\sigma), \tau, R; w\Big)\phi(s;\tau,\sigma)\hat{q}(\tau)dsd\tau - \kappa\mathcal{I}(\sigma) \tag{8}$$

where $\tilde{\tau}(s,\sigma) \equiv E_{\tilde{q}(\tau|s;\sigma)}[\tau]$ is the expected tax rate once the signal is observed and henceforth referred to as the perceived tax rate. In the following, we denote $f_{\tilde{\tau}}(\tau|\tau_0, w)$ the posterior distribution of $\tilde{\tau}(s,\sigma^\star)$ for a taxpayer with productivity w and signal s drawn from the Gaussian distribution with mean τ_0 and variance $\sigma^\star$.

3.3 Tractable Gaussian model

The Bayesian learning model presented above is generally intractable. To gain further insights, we focus here on a tractable formulation in order to highlight some predictions and implications.

As highlighted in the inattention literature (e.g. Maćkowiak and Wiederholt (2015); Mackowiak et al. (2018)), a closed form solution to problem (8) can be obtained under a Gaussian assumption.

Assumption 3.5 (tractable Gaussian learning)**.** *Let the prior $\hat{q}(\tau)$ be the Gaussian distribution*[11] *with mean $\hat{\tau}$ and variance $\hat{\sigma}^2$ and assume that agents use a first order approximation of their indirect utility to choose their attention strategies.*

As is demonstrated in Appendix A.2, the solution to this problem takes a simple form under Assumptions 3.1-3.5. First, the perceived tax rate $\tilde{\tau}$ is given by a linear combination of the prior mean $\hat{\tau}$ and the realization of the signal s

$$\tilde{\tau}(s,\sigma^\star) = \xi(\sigma^\star)s + (1-\xi(\sigma^\star))\hat{\tau} \tag{9}$$

where $\xi(\sigma^\star) \equiv \frac{\hat{\sigma}}{\hat{\sigma}+\sigma^\star} \in [0,1]$ is a measure of attention strategies. Therefore, the lower ξ is, the more taxpayers rely on their prior mean $\hat{\tau}$ and the less attention they devote to acquiring information.

Second, the optimal attention strategy ξ is given by

$$\xi(\sigma^\star) = \max\left(0, 1 - \frac{\kappa}{2\hat{\sigma}^2 \int \frac{\partial y^\star(\tilde{\tau};w)}{\partial(1-\tilde{\tau})}\Big|_{\tilde{\tau}=\tau}\phi(\tau;\hat{\tau},\hat{\sigma})d\tau}\right). \tag{10}$$

[11]One may instead consider that the prior is a truncated Gaussian with support on $[0,1]$ in order to ensure that the perceived tax rates $\tilde{\tau}$ always remain between zero and one. Doing so, the problem remains tractable but formulas become a lot less transparent. In practice, our simulations suggest that when the prior is sufficiently informative and the tax rate not too extreme, the posterior support belongs to $[0,1]$ a.s. with a gaussian prior.

Attention increases with the uncertainty in the prior $\hat{\sigma}^2$ and with productivity w. Indeed, more productive agents are more responsive to variations in the tax rate and therefore are willing to gather more information. Moreover, attention decreases with the cost to acquire and process information κ.

Predictions and implications: Attention strategies maximize the expected indirect utility. Because the latter is unaffected by a change in the realized tax rate $d\tau_0$, so are attention strategies. As a result, a variation in the tax rate τ_0 induces a change in the distribution of the posterior $f_{\tilde{\tau}}(\tilde{\tau}|w)$ only through a change in the signal s which is now drawn from a novel distribution $\phi(s; \tau_0 + d\tau_0, \sigma^\star)$. Therefore, taking the prior as given, taxpayers' perceived tax rate slowly adjusts to news and perceptions are somewhat anchored on the prior. Anchoring is a widely documented bias in the behavioral literature (see Gabaix (2019)). In our context, it has two major implications.

First, if agents' prior about taxes is biased ($\hat{\tau} \neq \tau_0$) and agents are not fully attentive ($\xi < 1$), the posterior is biased as well. In the tractable Gaussian case, we have

$$\tau_0 - E_{\phi(s)}[\tilde{\tau}(s, \sigma^\star)] = \tau_0 - [\xi\tau_0 + (1-\xi)\hat{\tau}] = (1-\xi)(\tau_0 - \hat{\tau}) \tag{11}$$

We will refer to this property as the *bias effect*.

Second, taxpayers labor supply responses to changes in the tax rate are attenuated by anchoring. Indeed, taking the prior as given, responses to tax changes only transit through the variation of the signal which is weighted by attention ξ. To illustrate this property, consider a first-order approximation to agents w aggregate labor supply $Y^\star_w \equiv \int_w y^\star(\tilde{\tau}; w) f_{\tilde{\tau}}(\tau|\tau_0, w) d\tilde{\tau}$ and its response to a tax change:

$$Y^\star_w \approx y^\star\Big(\xi\tau_0 + (1-\xi)\hat{\tau}; w\Big) \implies \frac{dY^\star_w}{d\tau_0} \approx \xi \frac{dy^\star}{d\tau} \tag{12}$$

Intuitively, $\frac{dy^\star}{d\tilde{\tau}}$ relates to preferences and how agents react to changes they perceive while ξ is a dampening factor that captures the fraction of the tax change that is perceived. As a result the elasticity of aggregate labor supply with respect to changes in the actual net-of-tax rate decreases in the presence of inattention which is why we refer to this property as the *elasticity effect*.

A third effect stems from the noisiness of the signal. Because agents observe different realizations of the signal s, it generates dispersion in perceived tax rates $\tilde{\tau}$ among agents of type w. This *dispersion effect* is captured in the distribution $f_{\tilde{\tau}}(\tau|\tau_0, w)$ and negatively affects the allocative efficiency of the economy. To see it formally, consider a second-order approximation of aggregate labor supply $Y^\star_w$:

$$Y^\star_w \approx y^\star\Big(E_{\phi(s)}[\tilde{\tau}]; w\Big) + \frac{1}{2}\frac{d^2 y^\star(\tilde{\tau}; w)}{d\tilde{\tau}^2}\Big|_{E_{\phi(s)}[\tilde{\tau}]} V_{\phi(s)}[\tilde{\tau}] \tag{13}$$

The first-term in the approximation relates to the *bias effect.* As before, taking the situation without biases as a counterfactual, the bias effect increases the allocative efficiency of the economy if and only if individuals understimate the tax rate. The second-term relates to the *dispersion effect.* It always has a negative impact on aggregate labor supply through the concavity of $y^{\star}(\tilde{\tau}; w)$. The intuition behind these efficiency losses is that agents with high perceived tax rates reduce their work effort significantly more than agents with low perceived tax rates increase theirs.

The structure of the model and its predictions seem consistent with the bulk of the empirical evidence on tax misperceptions. First, the bias effect enables us to account for any systematic bias in perceptions. For instance, the presence of an ironing heuristic (Rees-Jones and Taubinsky, 2016) or a salience bias (Chetty et al., 2009) leading to the underestimation of taxes would naturally arise with a prior featuring a downward bias. At the same time, the endogenous information acquisition component of the model is consistent with the recent evidence that these biases decrease when agents face higher tax rates (Taubinsky and Rees-Jones, 2017). Second, the elasticity effect is consistent with the lower elasticities measured in these contexts. Last, the model also accounts for dispersion which has been shown to be quantitatively important for the welfare implications of misperceptions (Taubinsky and Rees-Jones, 2017).

4 Positive theory of tax policy

This section analyzes the problem of the government and highlights its incentive to increase the tax rate in the presence of information frictions. In equilibrium – when the prior is adjusted – taxpayers anticipate the government incentive to deviate and the equilibrium tax rate is indeed higher.

4.1 Equilibrium tax policy

Sequential game We consider a welfarist government who maximizes a general social welfare function that sums an increasing and weakly concave transformation $G(.)$ of taxpayers' indirect utilities net of their information costs. It chooses a target tax schedule (τ_g, R_g), where τ_g is the marginal tax rate and R_g the demogrant, taking the distribution of skills $f_w(w)$ and agents common prior $\hat{q}(\tau)$ as given.

Having emphasized earlier that the prior distribution $\hat{q}(\tau)$ accounts for policy uncertainty, we here introduce uncertainty in the model by assuming that the government ability to implement its target tax schedule (τ_g, R_g) is uncertain. For instance, tax reforms may be the outcome of an uncertain political game in which the government may propose a reform that will ultimately be amended by the opposition or the parliament. Let ϑ be an implementation shock on the government target tax rate τ_g, namely the uncertainty surrounding the government's ability to implement its preferred tax rate[12]. Then, the true realized tax rate is $\tau_0 = \tau_g + \vartheta$ where ϑ is a white noise drawn from a known distribution $f_\Theta(\vartheta)$ which does not depend on the target tax rate τ_g. Since we impose that the government budget constraint is always satisfied, the actual demogrant R_0 adjusts to the realization of the implementation shock ϑ.

Formally, we model tax policy as the solution to the following sequential game

1. Agents are endowed with a prior $\hat{q}(\tau)$ and choose their attention strategies.

2. The government observes the prior $\hat{q}(\tau)$ and the distribution of skills $f_w(w)$ and sets the target tax policy (τ_g, R_g) that maximizes its objective under a resource constraint.

3. The actual tax rate $\tau_0 = \tau_g + \vartheta$ is implemented up to an implementation noise drawn from a known distribution $f_\Theta(\vartheta)$ and the actual demogrant R_0 adjusts to the resource constraint.

4. Taxpayers observe a Gaussian signal s about τ_0 which precision depends on their attention strategies and decide how much to consume and earn.

5. The government levies taxes and redistributes through the demogrant.

Tax policy The government's problem writes

$$\max_{\tau_g, R_g} \quad E_\vartheta\Big[\iint G\Big(\mathcal{V}(\tilde{\tau}, \tau_0, R; \kappa, w)\Big)\ f_{\tilde{\tau}}(\tau|\tau_0; w)\ f_w(w) d\tau dw\Big] \tag{14}$$

$$\text{s.t.} \quad \iint \tau_0 y^\star(\tilde{\tau}; w)\ f_{\tilde{\tau}}(\tau|\tau_0; w)\ f_w(w) d\tau dw \geq R_0 + E \tag{15}$$

where E is an exogenous expenditure requirement and $f_{\tilde{\tau}}(\tau|\tau_0; w)$ the posterior distribution of perceived rates for a taxpayer with productivity w when the actual tax rate is $\tau_0 = \tau_g + \vartheta$.

We characterize the optimal policy chosen by the government in the following proposition where first-order conditions have to hold in expectation because, from the government's perspective, its capacity to implement its desired policy (τ_g, R_g) is uncertain.

[12]We need a source of uncertainty to formally close the model and induce taxpayers to acquire information in this context but whether we consider implementation shocks or another form of ex ante uncertainty is not fundamental and would not affect our results (see Matějka and Tabellini (2017) for a related discussion).

Proposition 1. *Under assumptions 3.1-3.4, government's tax policy* (τ_g, R_g) *solves*

$$(\tau_g): \quad E_\vartheta\Bigg[\int\Big\{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)d\tau + \int\Big[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\Big]\frac{df_{\tilde{\tau}}(\tau|\tau_0;w)}{d\tau_g}\Big|_{\hat{q}(.)}d\tau\Big\}f_w(w)dw\Bigg] = 0 \tag{16}$$

$$(R_g): \quad E_\vartheta\Bigg[\iint\Big[\frac{G'(\mathcal{V})}{p} - 1\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)f_w(w)d\tau dw\Bigg] = 0 \tag{17}$$

together with the resource constraint (15) and where p represents the social marginal cost of public funds.

Proof. See appendix A.3 □

The first order condition (16) captures the two elements associated to a marginal increase in the tax rate. The first line measures the impact of the reform on allocations when the distribution of perceptions remains fixed i.e. absent behavioral responses. It corresponds to the standard mechanical and welfare effects (e.g. Saez (2001)): a marginal increase in the tax rate mechanically increases tax revenue by $y^\star d\tau_0$ additional dollars but reduces taxpayers' consumption and thus welfare by $\frac{G'(\mathcal{V})}{p}y^\star\ d\tau_0$ dollars.

The second line in condition (16) relates to the impact of the reform on the distribution of perceptions and thus captures behavioral responses to the reform. Indeed, behavioral responses transit through variations in the posterior distribution $f_{\tilde{\tau}}(\tau|\tau_0, w)$ of perceived tax rates $\tilde{\tau}$ which reflect changes in the actual tax rate τ_0. A marginal increase in the tax rate increases (on average) the perceived tax rate $\tilde{\tau}$ and thus reduces tax revenue by $\tau_0 y^\star(\tilde{\tau})d\tilde{\tau}$. This is a reformulation of the standard behavioral effect (e.g. Saez (2001)) in this context. Moreover, because agents misoptimize, the envelope theorem no longer applies and a marginal deviation from taxpayers's perceived rate induces a welfare cost equal to $\frac{G(\mathcal{V}(\tilde{\tau}))}{p}d\tilde{\tau}$. This new welfare effect introduces a corrective motive for taxation in the presence of misperceptions.

Condition (17) simply states that marginal social welfare weights $g \equiv \frac{G'(\mathcal{V})}{p}$ average to 1 at the optimum: the government is indifferent between having an additional dollar or redistributing an additional dollar.

Equilibrium Having characterized the optimal policy chosen by the government, we now turn to the definition of the equilibrium. An equilibrium is a set of target tax policy $(\tau_g^{\text{eq}}, R_g^{\text{eq}})$ and attention, consumption and earnings strategies solving the above sequential game. Importantly, in equilibrium the agents' common prior $\hat{q}(\tau)$ is consistent with the government target tax rate and with the uncertainty induced by the implementation shock. To account for the findings from the empirical literature presented before, we allow in equilibrium for a potential perception bias b in agents' common prior.

Given the structure of the problem, the only free variables are agents' common prior $\hat{q}(\tau)$, their attention strategies and the government target tax rate τ_g. Once they are simultaneously identified, remaining allocations (e.g. consumption and earnings) may be directly deduced. Hence, for the sake of simplicity and clarity, our formal definition of the equilibrium only involves these strategic variables.

Definition 4.1 (Positive equilibrium)**.** *Given the distribution of the implementation shock $f_\Theta(\vartheta)$, the equilibrium is a set of a target tax rate τ_g^{eq} chosen by the government and attention strategies chosen by the agents such that*

(a) Attention strategies solve agents' problem (8) given a prior distribution $\hat{q}(\tau)$.

(b) The target tax rate $\tau_g^{\mathrm{eq}} \in [0, 1]$ solves the government problem (14) given a common prior distribution $\hat{q}(\tau)$.

(c) The common prior distribution $\hat{q}(\tau)$ is the pdf of $\tau_g^{\mathrm{eq}} + b + \vartheta$.

Condition (c) implies that the prior is adjusted in equilibrium.[13] In particular, we have $E_{\hat{q}(\tau)}[\tau] = \tau_g^{\mathrm{eq}} + b$ in equilibrium. Consequently, when taxpayers correctly anticipate the government policy in equilibrium ($b = 0$), they nevertheless remain uncertain about the government capacity to implement its target policy as they anticipate that there will be an implementation shock ϑ. Their attention strategy then reflects their willingness to observe the implementation shock ϑ, which is indeed the only information conveyed through the signals.

4.2 Equilibrium tax policy with Gaussian shocks

This section characterizes the equilibrium tax policy under Assumptions 3.1-3.5 and Gaussian implementation shocks. Indeed, with ϑ a Gaussian white noise of variance σ_ϑ^2, we have that the equilibrium distribution of the common prior is also Gaussian. Hence, the equilibrium distribution of the posterior $f_{\tilde{\tau}}(\tau|\tau_0, w)$ is Gaussian with mean $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$.

Using a first-order approximation of the integrands, we show in Appendix A.4 that the optimality conditions characterizing the government tax policy (Proposition 1) are in equilibrium equal to

$$E_\vartheta\left[\int \left\{(1-g)\,y^\star + \Big(g(1-\xi)(b-\vartheta) + \tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\xi\right\}\Big|_{\tilde{\tau}=\mu} dF_w(w)\right] = 0 \qquad (18)$$

$$E_\vartheta\left[\int g_{|\tilde{\tau}=\mu}\, dF(w)\right] = 1 \qquad (19)$$

[13] For the equilibrium to exist, it requires that $P(\tau_g^{\mathrm{eq}} + b + \vartheta \notin [0, 1]) = 0$.

with social marginal welfare weights $g \equiv \frac{G'(\mathcal{V})}{p}$. Equation (18) provide a transparent illustration of taxpayers' behavioral response to tax changes and their impact for the optimal tax policy. Taxpayers here respond to perceived tax changes, behavioral responses $\frac{dy^\star}{d\tilde{\tau}}$ are thus attenuated by attention ξ and an increase in the tax rate changes the government revenue by $\frac{dy^\star}{d\tilde{\tau}}\xi y^\star$. It exactly corresponds to the aforementioned *elasticity effect* and it pushes for higher tax rates.

Moreover, the new welfare effect is here clearly associated with a failure of the envelope theorem and the aforementioned *utility misoptimization costs.* It is exactly proportional to the size of the bias in the posteriors $\mu - \tau_0 = (1-\xi)(b-\vartheta)$ which captures the magnitude of these utility costs. It generates a corrective motive for taxation and calls on average for higher tax rates when agents underestimate taxes ($b < 0$) and lower tax rates when agents overestimate taxes ($b > 0$) as illustrated Figure 3 below.

We further provide a simple and transparent formula for the equilibrium tax rate expressed in terms of sufficient statistics. The first term corresponds to the textbook optimal linear tax formula (e.g. Piketty and Saez (2013)) up to the *elasticity effect* captured by the presence of attention ξ and the second term corresponds to the corrective motive associated with *utility misoptimization costs* in the presence of a perception bias b. We show in Appendix A.5 that it writes

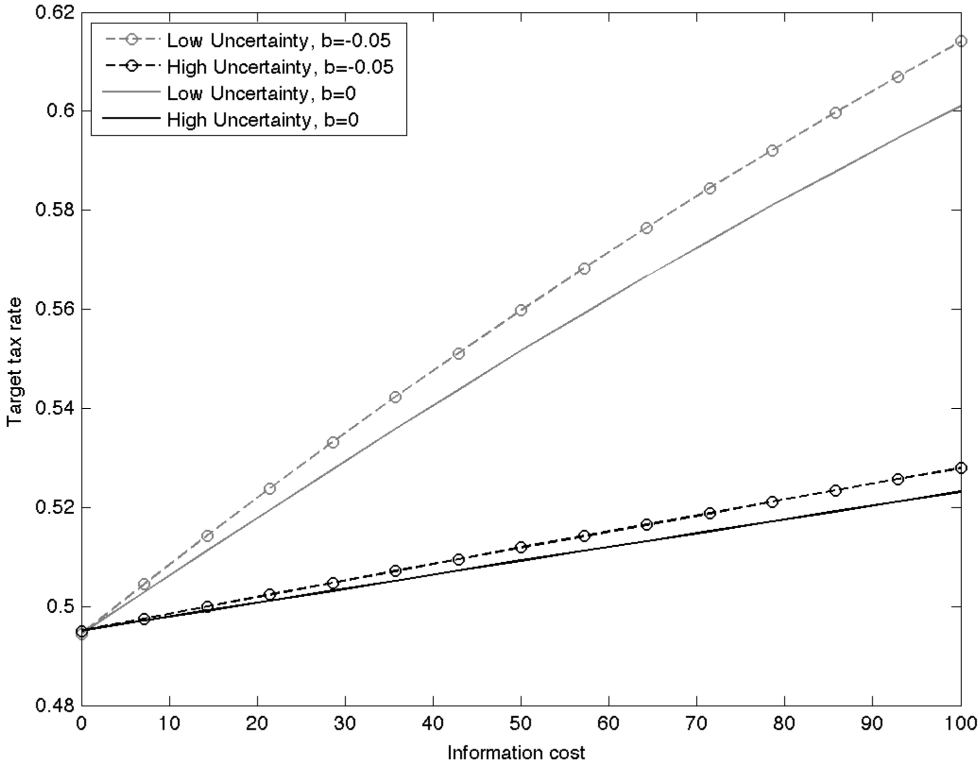
$$\tau_g^{\text{eq}} = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\ \frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{20}$$

where e is the labor supply elasticity with respect to the perceived marginal net of tax rate assumed constant (e.g. iso-elastic utility function) and where we have introduced the mean operator $\bar{x} = \int x(w)f(w)dw$. This formula relies on small noise ϑ and small perception bias b first-order approximations and all endogenous quantities on the right hand side of the equation are evaluated at τ_g^{eq}.

Figure 3 plots the equilibrium tax rate implemented by the government as a function of agents' information cost parameter κ. Lower information costs or higher uncertainty increases taxpayers' attention and thereby taxpayers' responsiveness to changes in the tax rate. In equilibrium, the tax rate is then closer to its counterpart with full information: they limit the government's ability to fool taxpayers and thus the incentive for the government to deviate towards higher tax rates.

Simulations reveal that when $\kappa = \$50/\text{bit}$ and priors are unbiased in equilibrium ($b = 0$), the median taxpayer spends \$51 (resp. \$17) per year on information acquisition when the implementation uncertainty is $\hat{\sigma}_\vartheta = 0.1$ (resp. 0.05). The equilibrium tax rate increases by 1 (resp. 6) percentage point and the average demogrant by \$430/year (resp. $\$1,581$/year) in comparison to a situation with perfect information. While these numbers

Figure 3: Target tax rate in equilibrium



Note: Equilibrium target tax rates for different values of the information cost κ expressed in $\$/bit/yr$. Low (resp. high) uncertainty corresponds to Gaussian implementation shocks with a standard deviation equal to 0.05 (resp. 0.1). b is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition 1. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

are only presented for illustrative purpose, they clearly indicate that small information rigidities may lead to large deviations in equilibrium.

5 Normative benchmark and the taxation bias

Having characterized the equilibrium tax policy, we now turn to the characterization of the optimal tax policy from a normative perspective. We show that in the presence of information frictions the equilibrium tax rate is substantially higher than the normative tax rate. We refer to this phenomenon as the taxation bias.

5.1 Normative tax policy and the taxation bias

The normative tax policy is the equilibrium tax policy that would be chosen by a benevolent social planner who has the same information as the government but accounts for the equilibrium effects of its policy. In other words, it corresponds to the policy which maximizes the equilibrium social welfare. Formally, the social planner solves the same maximization problem (14) as the government with the additional equilibrium restriction (c) from Definition 4.1. Intuitively, the social planner internalizes the impact of the choice of tax policy on agents' priors and attention strategies when making its choice. The normative tax policy is characterized by the following first order conditions.

Proposition 2. *Under Assumptions 3.1-3.4, social planner's tax policy* $(\tau_g^\star, R_g^\star)$ *solves*

$$\begin{aligned}(\tau_g^\star) \quad &: \quad E_\vartheta\Bigg[\int\Big\{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)d\tau \\ &\qquad + \int\Big[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\Big]\frac{df_{\tilde{\tau}}(\tau|\tau_0;w)}{d\tau_g}d\tau\Big\}f_w(w)dw\Bigg] = 0 \qquad (21) \\ (R_g^\star) \quad &: \quad E_\vartheta\Bigg[\iint\Big[\frac{G'(\mathcal{V})}{p} - 1\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)f_w(w)d\tau dw\Bigg] = 0 \qquad (22)\end{aligned}$$

together with the resource constraint (15) and where p represents the social marginal cost of public funds.

Proof. See Appendix A.3 □

As before, conditions have to hold in expectation because the social planner's ability to implement its target policy τ_g is uncertain. The only difference between Proposition 1 and Proposition 2 is that the derivative $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_0;w)}{d\tau_g}$ in equation (21) now reflects changes in both the signal received (information acquisition) and the agents' prior (equilibrium relation).

Definition 5.1 (taxation bias)**.** *We call taxation bias the difference between the positive tax rate* τ_g^{eq} *and the normative tax rate* $\tau_g^\star$*.*

Whenever tax revenue decreases more on average in response to an increase in the tax rate when one allows agents' prior to adjust,[14] then the taxation bias is positive. Furthermore, it is worth noting that the government cannot credibly commit to implement the normative tax rate given the existing information frictions as it would always have an incentive to deviate towards higher tax rates ex-post.

5.2 Analytical characterization with Gaussian shocks

To provide an analytical and tractable expression for the taxation bias we again turn to the Gaussian case. As before, under Assumptions 3.1 - 3.5 and Gaussian implementation shocks ϑ, the posterior distribution $f_{\tilde{\tau}}(\tau|\tau_0;w)$ is Gaussian with mean $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$.

[14] That is, when $\int \tau_0 y^\star \frac{df_{\tilde{\tau}}(\tau|\tau_0;w)}{d\tau_g}d\tau \leq \int \tau_0 y^\star \frac{df_{\tilde{\tau}}(\tau|\tau_0;w)}{d\tau_g}\big|_{\hat{q}(.)}d\tau$. A fairly weak condition.

Using a first-order approximation of the integrands, we show in Appendix A.4 that the optimality conditions that characterize the social planner's tax policy (Proposition 2) are in equilibrium equal to

$$E_\vartheta\left[\int \left\{(1-g)y^\star + \Big(g(1-\xi)(b-\vartheta)+\tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}(b-\vartheta)\Big)\right\}\Big|_{\tilde{\tau}=\mu} f_w(w)dw\right] = 0$$

$$E_\vartheta\left[\int g\big|_{s=\tau_g+\vartheta} f(w)dw\right] = 1 \tag{23}$$

Whereas taxpayers responses to a change in the tax rate are captured through $\frac{dy^\star}{d\tilde{\tau}}\xi$ in the positive equilibrium from equation (18), they are here captured through $\frac{dy^\star}{d\tilde{\tau}}\big(1-\frac{d\xi}{d\tau_g}(b-\vartheta)\big)$. This comes from the fact that, to a first-order, the government evaluates changes in $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$ taking the prior and thus attention ξ as given whereas the social planner internalizes its impact on both the prior mean $\hat{\tau}$ and attention ξ. As a result of the former, taxpayers perceived responsiveness is higher for the social planner than for the government: the *elasticity effect* disappears once the prior is adjusted. The latter is reflected in the additional term $\frac{d\xi}{d\tau_g}(b-\vartheta)$ which captures changes in attention ξ triggered by increases in the tax rate. Indeed, by shifting the support of the prior, increases in the tax rate change the points at which agents evaluate the gains from information and thus the value of acquiring information holding uncertainty and information costs constant.

We again derive a simple sufficient statistics formula for the normative tax rate which is the analogue to equation (20) for the positive tax rate. We show in Appendix A.5 that it writes

$$\tau_g^\star = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\ \frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{24}$$

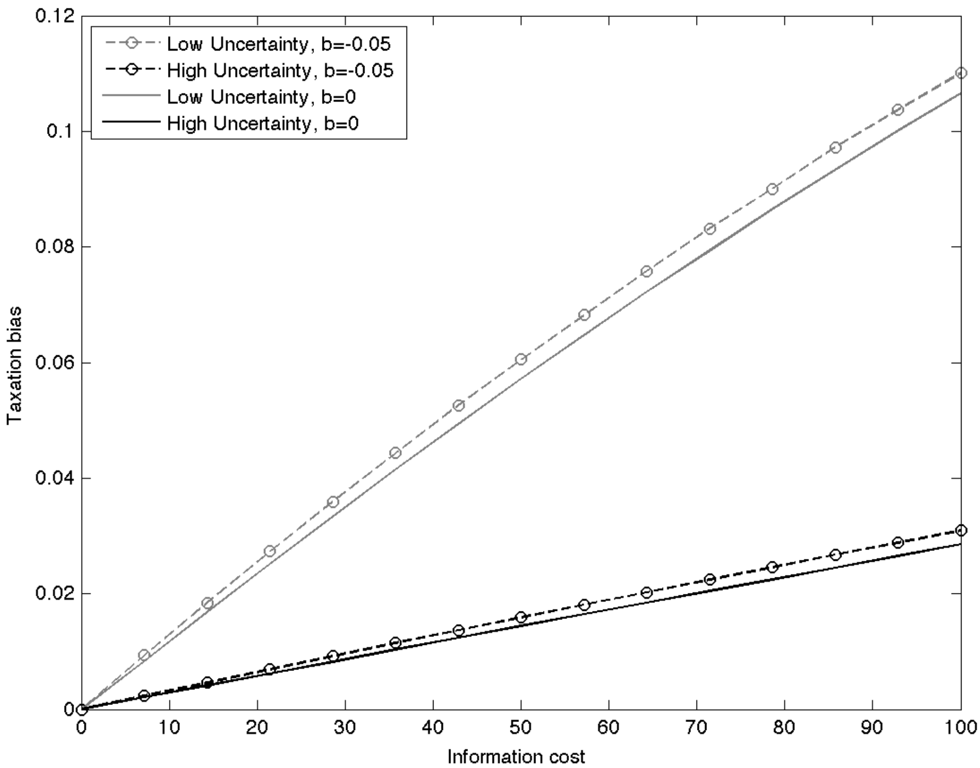
where all endogenous quantities on the right hand side are evaluated at $\tau_g^\star$. Now that the *elasticity effect* has disappeared, the first term exactly corresponds to the textbook optimal linear tax formula (e.g. Piketty and Saez (2013)) while the second term again corresponds the corrective motive associated with *utility misoptimization costs* in the presence of a perception bias b.

These corrective terms are not identical in equations (20) and (24) depending on the presence or absence of the elasticity effect. However, they have the same sign for both positive and normative tax rates and are always proportional to the perception bias b. As a result, they can be ignored in the analysis of the taxation bias for small perception biases (see also Figure 4). Since sufficient statistics formulas (20) and (24) are obtained under a small perception bias b first-order approximation, we use them to obtain a simple analytical formula for the taxation bias assuming corrective terms can be ignored to a first-order. We furthermore assume that the taxation bias is small such that all endogenous quantities are evaluated at tax rate $t \approx \tau_g^\star \approx \tau_g^{\text{eq}}$ to obtain

$$\tau_g^{\text{eq}} - \tau_g^\star = \frac{e\ t^2}{(1-g)y^\star}\left(\overline{y^\star} - \overline{y^\star \xi}\right) \tag{25}$$

The taxation bias is fundamentally driven by agents' inattention to the tax rate. Because agents are inattentive, the government has an incentive to increase the tax rate and will do so, failing to internalize the fact that in equilibrium agents' prior adjusts to its policy. Since, it is an income-weighted measure of attention that enters the formula, the taxation bias is increasing in the degree of regressivity in attention. Moreover, the taxation bias increases with the elasticity of labor supply e, the tax rate t and decreases with the government's redistributive tastes.

Figure 4: Taxation bias



NOTE: Difference between positive and normative tax rates for different values of the information cost κ expressed in annual \$ / bit. Low (resp. high) uncertainty corresponds to Gaussian implementation shocks with a standard deviation equal to 0.05 (resp. 0.1). b is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition 2. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y,w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

We illustrate in Figure 4 that the taxation bias increases with agents' inattention.

Indeed, lower policy uncertainty or higher information costs tend to reduce agents' attention while they increase the size of the taxation bias. Importantly, even small information costs may induce large deviations from the normative policy in equilibrium. For instance, when policy uncertainty is low and the benefits to attention relatively modest, annual information costs of just 50 \$/bit (resp. 100 \$/bit) generate an equilibrium tax rate of 55% (resp. 60 %) whereas the normative tax rate is in both cases equal to 50%. The taxation bias is thus equal to 5% (resp. 10 %)[15]. This has important welfare implications which we now turn to.

6 Welfare Implications

This section describes the welfare implications of information frictions. It first decomposes the overall variation in social welfare using the normative and positive equilibria studied previously. It then tries to quantify the relative importance of the different channels through which information rigidities ultimately affect taxpayers' utilities.

6.1 Information rigidities and social welfare

Let $SW^{\text{eq}}(b,\kappa)$ be the social welfare from equation (14) evaluated at the positive equilibrium. The welfare impact of information rigidities is therefore $\Delta SW^{\text{eq}}(b,\kappa) \equiv SW^{\text{eq}}(b,\kappa) - SW^{\text{eq}}(0,0)$.[16] It may be decomposed between the potential gains from information rigidities and the welfare cost induced by the taxation bias as follows

$$\Delta SW^{\text{eq}}(b,\kappa) = \underbrace{SW^{\text{eq}}(b,\kappa) - SW^{\star}(b,\kappa)}_{\text{Taxation bias } (\leq 0)} + \underbrace{SW^{\star}(b,\kappa) - SW^{\star}(0,0)}_{\text{Potential gain}} \tag{26}$$

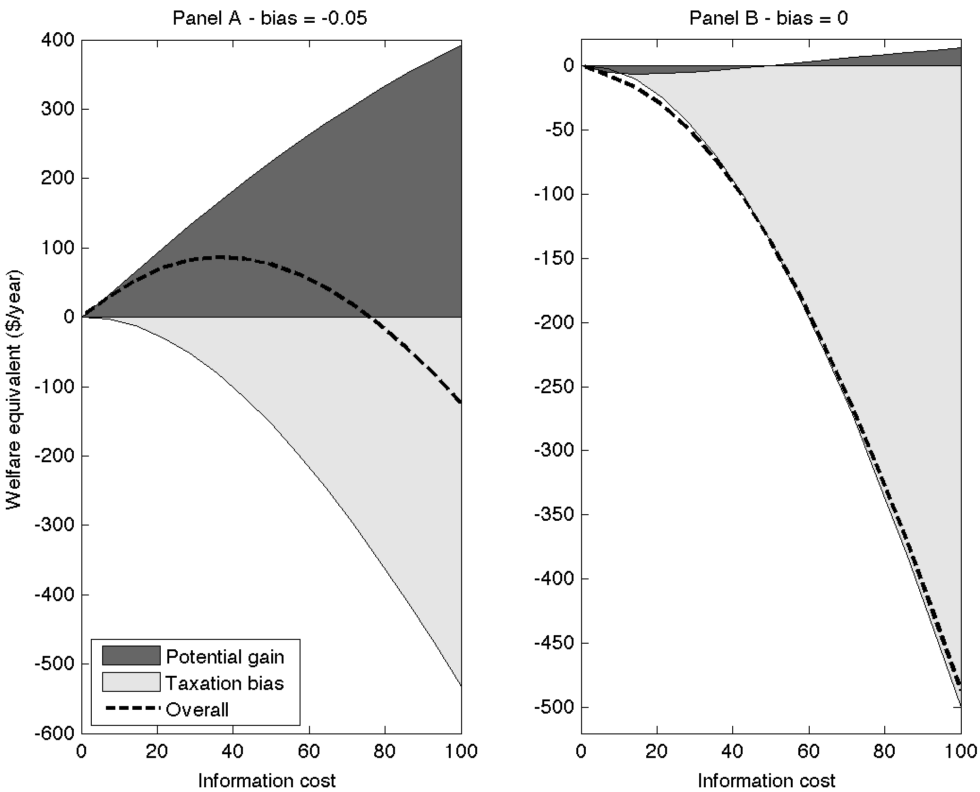
where $SW^{\star}(b,\kappa)$ is the social welfare evaluated at the normative equilibrium. The welfare impact of the taxation bias is then negative by definition of the normative equilibrium and information rigidities are welfare improving if and only if $SW^{\star}(b,\kappa) - SW^{\text{eq}}(0,0) \geq |SW^{\text{eq}}(b,\kappa) - SW^{\star}(b,\kappa)|$. This requires a downward bias in priors $b < 0$ such that agents underestimate tax rates. This can be easily seen when looking at our simple tax formula equation (24). The normative tax rate is equal to the optimal tax rate without information when $b = 0$, up to a first order approximation. Consequently, there is no first order gain from information rigidities in this case. However, equation (20) shows that the taxation bias generates a relatively large increase in the tax rate, so that the overall welfare cost is negative.

[15]Note that we do verify in our simulations that the impact of small perception biases b seems to be second-order in determining the size of the taxation bias.

[16]Note that $SW^{\text{eq}}(0,b) = SW^{\text{eq}}(0,0)$ since as soon as the information cost κ is nil, agents have perfect information and whether priors are biased is irrelevant.

Now, when taxpayers' expectations are downward biased in equilibrium, information rigidities can be potentially welfare increasing depending on the information cost κ. Intuitively, start from a situation such that the downward bias is large enough ($b < b^u$) to overcome the taxation bias and other negative welfare effects of misperceptions[17] when the information cost tends to zero.[18] Then, information rigidities are, by definition, welfare improving when the information cost is sufficiently small. However, as the information cost grows, the negative welfare effect induced by the taxation bias will generally increase more rapidly than the potential gain from the negative bias. Consequently, there might exist a large enough κ^u such that, for a given bias ($b < b^u$), information rigidities are welfare improving if and only if $\kappa \leq \kappa^u$.

Figure 5: Welfare decomposition



NOTE: Welfare decomposition from equation (26) for different values of the information cost κ. The standard deviation for the Gaussian implementation shocks is equal to 0.05. b is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition 1. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

[17] These stem from the signal distribution (*dispersion effect* and increased variance in allocations) and misoptimization costs associated to the implementation shocks. Simulations indicate that with Gaussian shocks, b^u is approximately equal -1.5 (resp. -2.5) percentage points when $\sigma_\vartheta = 0.05$ (resp.0.1).

[18] i.e. b^u is such that $\lim_{k\mapsto 0^+}(SW^{\text{eq}}(b^u, k) - SW^\star(0,0))/k = 0$

Figure 5 illustrates these mechanisms with Gaussian implementation shocks. When the downward equilibrium perception bias is equal to 5 percentage points (Panel A), the potential gain induced by information rigidities overcome the cost from the taxation bias as long as the annual information cost κ is lower than \$80 per bit. Above this threshold, the government deviation from the normative policy becomes too important and information rigidities are welfare decreasing. Panel B indicates[19] that information rigidities systematically impair welfare in the absence of perception bias in equilibrium. Consequently, a downward bias is a necessary, but not a sufficient, condition for tax misperceptions to be welfare improving.

6.2 Redistributive effect

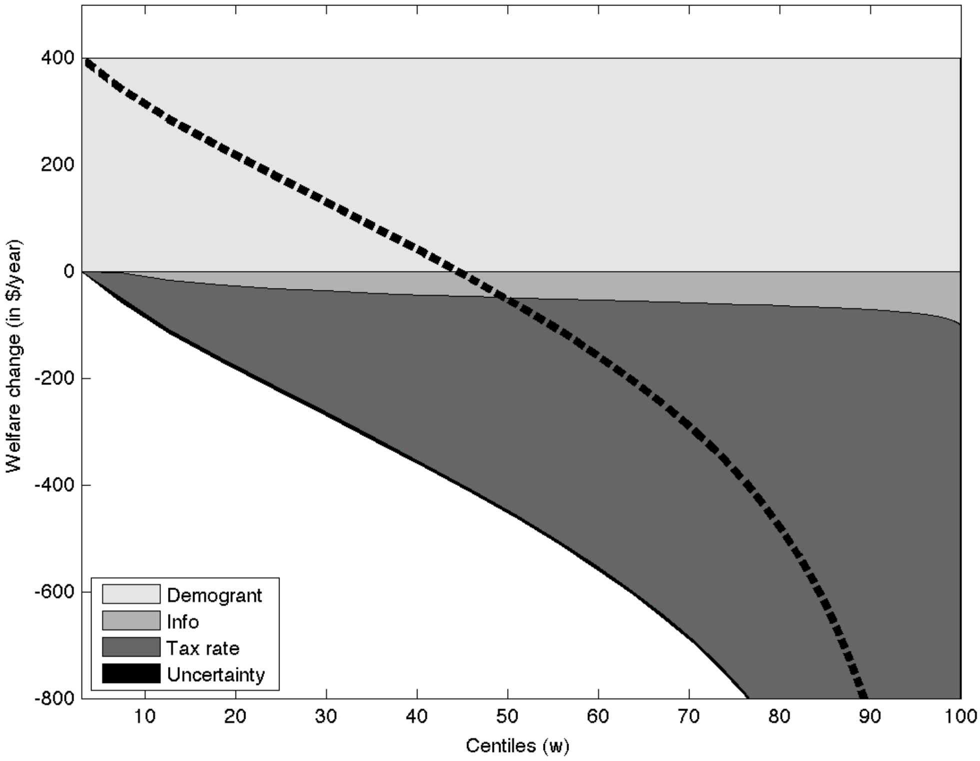
We now turn to an analysis of the welfare implications of information rigidities at the individual level. To this end, define $\Delta\mathcal{V} \equiv \mathcal{V}_\kappa(\tau_g^\star(\kappa), w) - \mathcal{V}_0(\tau_g^\star(0), w)$ the variation in the expected utility of a taxpayer with skill w evaluated at the positive equilibrium when the information cost is κ and a counterfactual with perfect information. The quasi-linear utility Assumption (3.3) allows to easily decompose the variation in expected utility induced by information rigidities in the following way (see Appendix A.7),

$$\Delta\mathcal{V} \quad = \quad \Delta^{\mathcal{V}} R + \Delta^{\mathcal{V}} \tau_g + \Delta^{\mathcal{V}} \text{info cost} + \Delta^{\mathcal{V}} b + \Delta^{\mathcal{V}} \text{uncertainty} \tag{27}$$

that is, the welfare impact of information rigidities at the individual level arises from a variation the demogrant R, the tax rate τ_g, the cost of acquired information $\kappa\mathcal{I}(\sigma^\star)$, the misoptimization costs induced by potential perception bias b, and the change in overall uncertainty. It is straightforward to show that in equilibrium the variation in the demogrant has a positive effect on expected utility while all other terms are negative. Also, information rigidities cannot Pareto improving in equilibrium and there must be winners and losers.

[19]Strangely enough, the potential gain is first decreasing and then increasing when b is small or nil. While the magnitude of the potential gain is small and thus negligible in comparaison to the impact of the taxation bias, it deserves to be briefly explained. Consider the two extreme cases where $\kappa = 0$ and $\kappa \mapsto \infty$. Hence, ξ is respectively equal to one or zero for each taxpayer. Everything else being equal, aggregate earnings are larger when $\kappa \mapsto \infty$ as agents behave as if there were no implementation shocks when deciding how much to earn (individual earnings are a concave function of the perceived rate), while they fully adjust to these shocks when $\kappa = 0$. Consequently, the potential gain converges to a positive value as κ tends to infinity. However, when κ is small but strictly positive, some taxpayers noisily observe the implementation shocks so that the variance of their earning choices increases. Ultimately, it lowers aggregate earnings and generates a negative and decreasing potential gain for small values of κ. Simulations indicate that the above described variations in aggregate earnings dominate other second order effects such as misoptimization costs here.

Figure 6: Variation in expected utility



NOTE: Expected utility decomposition from equation (27) by deciles of the productivity level w. The standard deviation for the Gaussian implementation shocks is equal to 0.1. There is no perception bias in agents' prior $b = 0$ and the information cost is $\kappa = 50$. The government has a log social welfare function and its policy follows from Proposition 1. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y,w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

Figure 6 plots the above expected utility decomposition with Gaussian implementation shocks ($\sigma_\vartheta = 0.1$), no perception bias in equilibrium ($b = 0$) and an information cost κ of \$50 per year. The increase in uncertainty has a negligible impact for all levels w. Moreover, and while it is not reported on this graph for clarity, the same conclusion holds for the misoptimization costs arising from a potential bias in equilibrium. Note that this validates the use of first order approximations of taxpayers' expected utility to derive optimal tax formulas in previous sections.

The redistributive impact of information rigidities is driven by the government incentive to increase the tax rate and thus redistribution through the demogrant. This policy naturally benefits low skill workers at the extent of high skill workers. These high skill workers suffer an additional penalty arising from the information cost. Indeed, in equilibrium, they are more willing to collect information in order to form an accurate forecast of the tax policy. Overall information costs represents a deadweight lost for society which is bore mostly by those with the highest utilities, however they remain relatively small in comparison to their large impact on tax policy and welfare.

Conclusion

The paper develops a positive theory of taxation with information frictions and shows that it leads to a taxation bias. Given the observed misperceptions, the gains associated to the normative equilibrium may be important enough to identify implementable solutions to the taxation bias. For central bankers, the preferred solution was the delegation of the monetary policy to an independent institution. Here, the taxation bias does not arise because of an arbitrage – e.g. between inflation and unemployment for central bankers – between two interconnected objectives. Therefore, delegation is not the solution. To this end, it would be interesting for future research to investigate the factors that generate the information cost κ. If the latter could be interpreted as a complexity measure of the tax system, then ensuring that the tax schedule is not too complex will restrain the government incentive to deviate from the normative policy.

While some of the results may be model-specific, a general lesson emerges for the analysis of the welfare consequences of information frictions. It should be clear that measuring downward biases in tax perceptions do not necessarily imply welfare improvements. They do lower the efficiency costs of taxation in existing tax systems, but existing tax systems without misperceptions are arguably not the right counterfactual to use for welfare analysis. Indeed, there may be other equilibrium effects at play – which take the form of a taxation bias in this paper. As we show, the welfare consequences of such equilibrium effects may be the dominant effect, overturning the previous welfare implications. One should thus be very careful with welfare implications drawn from the measurement of misperceptions. We believe that this general lesson applies outside of the realm of taxation.

References

Barro, R. J. and Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. *Journal of political economy*, 91(4):589–610. 3, 5, 8

Benabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *The Quarterly journal of economics*, 121(2):699–746. 12

Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. Technical report, National Bureau of Economic Research. 4

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018. 7, 20, 23, 25, 27

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5):1–33. 4

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–77. 2, 5, 6, 12, 15

Farhi, E. and Gabaix, X. (2015). Optimal taxation with behavioral agents. Technical report, National Bureau of Economic Research. 4, 6

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710. 2

Gabaix, X. (2019). *Behavioral Inattention.* Elsevier. 5, 12, 14

Gerritsen, A. (2016). Optimal taxation when people do not maximize well-being. *Journal of Public Economics*, 144:122–139. 4

Liebman, J. B. and Zeckhauser, R. J. (2004). Schmeduling. Technical report, Working Paper. 2, 5, 6

Lockwood, B. B. (2016). Optimal income taxation with present bias. Technical report, University of Pennsylvania Working Paper. 4

Maćkowiak, B. and Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *The Review of Economic Studies*, 82(4):1502–1532. 13

Mackowiak, B. A., Matejka, F., Wiederholt, M., et al. (2018). Survey: Rational inattention, a disciplined behavioral model. Technical report, CEPR Discussion Papers. 5, 13, 31

Matějka, F. and Tabellini, G. (2017). Electoral competition with rationally inattentive voters. *Forthcoming.* 3, 16

Miller, B. M., Mumford, K., et al. (2015). The salience of complex tax changes: evidence from the child and dependent care credit expansion. *National Tax Journal*, 68(3):477–510. 5

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208. 4

Piketty, T. and Saez, E. (2013). Chapter 7 - optimal labor income taxation. In Alan J. Auerbach, Raj Chetty, M. F. and Saez, E., editors, *Handbook of Public Economics, vol. 5*, volume 5 of *Handbook of Public Economics*, pages 391 – 474. Elsevier. 6, 9, 19, 22, 36

Reck, D. (2016). Taxes and mistakes: What's in a sufficient statistic? 4

Rees-Jones, A. and Taubinsky, D. (2016). Heuristic perceptions of the income tax: Evidence and implications for debiasing. Technical report, National Bureau of Economic Research. 2, 5, 9, 12, 15

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–229. 4, 17, 32

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690. 5, 6, 12

Taubinsky, D. and Rees-Jones, A. (2017). Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4):2462–2496. 5, 15

A Appendix

A.1 Solution to the Toy Model with Imperfect Information

The government seeks to maximize tax revenue taking the prior as given. Its problem writes $\max_\tau \tau Y(1-\tilde{\tau})$ such that $\tilde{\tau} = \xi\tau + (1-\xi)\hat{\tau}$ and $\tau \in [0,1]$. The associated Lagrangian is $\mathscr{L}(\tau,\lambda) = \tau Y(1-\xi\tau-(1-\xi)\hat{\tau}) + \lambda(\tau - 1)$. Following from the first order Kuhn and Tucker conditions, $\tau = 1$ if and only if $\hat{\tau} \leq 1 - \frac{\xi}{1-\xi}e$ and $\tau = \frac{1}{1+\xi e}$ otherwise. These conditions are also sufficient since the problem is convex under the assumption that $\tau Y(1-\tau)$ is concave.

At the rational equilibrium, the prior is correct $\hat{\tau} = \tau$. Guess that the rational equilibrium is interior. Hence, $\hat{\tau} = \frac{1}{1+\xi e}$. Because $e > 0$, it implies that $\hat{\tau} > 1 - \frac{\xi}{1-\xi}e$ thus confirming the guess. The optimal policy of the government at the rational equilibrium is therefore to set $\tau^\star = \frac{1}{1+\xi e}$. It is then straightforward to prove that $\tau^\star Y(1-\tau^\star) < \tau^n Y(1-\tau^n)$ where $\tau^n \equiv \frac{1}{1+e}$ as $\tau^n = \arg\max_{\tau\in[0,1]} \tau Y(1-\tau)$. Moreover, the taxation bias $\tau^\star - \tau^n = \frac{(1-\xi)e}{(1+\xi e)(1+e)}$ is strictly positive for all $\xi \in (0,1)$.

A.2 Reformulation of the Tractable Gaussian Learning

Assumptions 3.1-3.4 imply that the indirect utility of a taxpayer is given by equation (4). Performing a first order Taylor approximation of the latter around τ_0 gives

$$V^2_{\tau_0}(\tilde{\tau}, \tau_0, R_0; w) = V(\tau_0, \tau_0, R_0; w) + (\tilde{\tau} - \tau_0)\frac{\partial V}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau_0} \tag{A.1}$$

where $\frac{\partial V}{\partial \tilde{\tau}} = (\tilde{\tau} - \tau_0)\frac{\partial y^\star}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau_0}$ from (3). Hence,

$$\iint V^2_\tau(\tilde{\tau}, \tau, R; w)\phi(s;\tau,\sigma)\phi(\tau;\hat{\tau},\hat{\sigma})ds d\tau = \int \left[V(\tau,\tau,R;w) + \tilde{\sigma}^2 \frac{\partial y^\star}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau}\right]\phi(\tau;\hat{\tau},\hat{\sigma})d\tau$$

where $\tilde{\sigma}^2$ is the posterior variance and we used the fact that with a gaussian prior and likelihood, the posterior is gaussian. Accordingly, the expected information reduction writes

$$\mathcal{I}(\sigma) = \frac{1}{2}\big(\log(2\pi e\hat{\sigma}^2) - \log(2\pi e\tilde{\sigma}^2)\big) = \frac{1}{2}\log\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \tag{A.2}$$

where $\frac{1}{2}\log(2\pi e\sigma^2)$ is the differential entropy (in bits) of a gaussian with variance σ^2. Therefore, under assumption 3.5, problem (8) becomes

$$\max_{\tilde{\sigma}\geq\hat{\sigma}} \quad \tilde{\sigma}^2 \int \frac{\partial y^\star}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau}\phi(\tau;\hat{\tau},\hat{\sigma})d\tau - \frac{\kappa}{2}\log\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \tag{A.3}$$

This problem has been extensively studied in the literature. For instance, a step-by-step derivation of the solution is provided in Mackowiak et al. (2018). It shows that the perceived tax rate is $\tilde{\tau} = \xi s + (1-\xi)\hat{\tau}$ where $\xi \in [0,1]$ is a measure of the attention level. At the optimum, a taxpayer chooses

$$\xi = \max\left(0, 1 + \frac{\kappa}{2\hat{\sigma}^2 \int \frac{\partial y^\star}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau} \phi(\tau; \hat{\tau}, \hat{\sigma}) d\tau}\right) \tag{A.4}$$

A.3 Proof of Proposition 1 and Proposition 2

We here prove both propositions at the same time since the only difference between the two problems is in the nature of responses to tax changes that are taken into account. We thus solve the general problem where all agents' responses are taken into account (including equilibrium relations) to obtain Proposition 2 and from which Proposition 1 naturally follows.

The Lagrangian associated to the general problem (14) writes

$$\begin{aligned} \mathscr{L}(\tau_g, R, p) &= E_\vartheta\Big[\iint \Big[G\Big(\mathcal{V}(\tilde{\tau}, \tau_g + \vartheta, R, \kappa; w)\Big) \\ &\quad + p\Big((\tau_g + \vartheta) y^\star(\tilde{\tau}; w) - R_0 - E\Big)\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) f_w(w) d\tilde{\tau} dw\Big] \end{aligned} \tag{A.5}$$

The derivation of the first order conditions follows from a perturbation approach (Saez (2001)). Consider a reform that marginally increases τ_g by $d\tau_g$. The impact of the reform on the objective is

$$\begin{aligned} \frac{1}{p}\frac{d\mathscr{L}}{d\tau_g} &= E_\vartheta\Big[\int \Big\{ \int \Big[\frac{G'(\mathcal{V})}{p}\frac{d\mathcal{V}}{d\tau_g} + y^\star\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) d\tilde{\tau} \\ &\quad + \int \Big[\frac{G(\mathcal{V})}{p} + (\tau_g + \vartheta) y^\star - R_0 - E\Big] \frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w)}{d\tau_g} d\tilde{\tau}\Big\} f_w(w) dw\Big] \end{aligned} \tag{A.6}$$

where $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}$ is the change in the posterior distribution of perceived tax rate for type w and captures agents' responses to tax changes.

By definition $\int f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) d\tilde{\tau} = 1$ thus $\int \frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g} d\tilde{\tau} = 0$. Also, by Assumption 3.3 we have $\frac{d\mathcal{V}}{d\tau_g} = -y^\star(\tilde{\tau}; w)$. Therefore, the optimality condition $\frac{1}{p}\frac{d\mathscr{L}}{d\tau_g} = 0$ writes

$$\begin{aligned} E_\vartheta\Big[\int \Big\{ \int \Big[-\frac{G'(\mathcal{V})}{p} y^\star + y^\star\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) d\tilde{\tau} \\ + \int \Big[\frac{G(\mathcal{V})}{p} + (\tau_g + \vartheta) y^\star\Big] \frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w)}{d\tau_g} d\tilde{\tau}\Big\} f_w(w) dw\Big] = 0 \end{aligned} \tag{A.7}$$

This is equation (21) from Proposition 2 and characterizes the optimal tax rate chosen by the social planner. Equation (16) from Proposition 1 which characterizes the optimal tax rate chosen by the government is obtained by when agents' responses to a change in the tax rate is computed holding agents' prior $\hat{q}$ constant. That is $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}\Big|_{\hat{q}(.)}$ replaces $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}$ in equation (A.7).

Now, consider a reform that consists in a marginal change dR of the demogrant. The impact of this reform on (A.5) is

$$\frac{1}{p}\frac{d\mathscr{L}}{dR} = E_\vartheta\Big[\iint \Big[\frac{G'(\mathcal{V})}{p}\frac{d\mathcal{V}}{dR} - 1\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) f_w(w) d\tilde{\tau} dw\Big] \tag{A.8}$$

From Assumption 3.3, $\frac{d\mathcal{V}}{dR} = 1$. The optimality condition $\frac{1}{p}\frac{d\mathscr{L}}{dR} = 0$ thus writes

$$E_\vartheta\Big[\iint \Big[\frac{G'(\mathcal{V})}{p} - 1\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g + \vartheta; w) f_w(w) d\tilde{\tau} dw\Big] = 0 \tag{A.9}$$

This is equation (17) from Proposition 1 and equation (22) from Proposition 2.

A.4 Optimal policies in tractable Gaussian case

Conditions (A.7) and (A.9) rely only on Assumptions 3.1-3.4 and therefore applies to any learning leading to a differentiable posterior distribution of perceptions $f_{\tilde{\tau}}(\tau|\tau_0; w)$ with positive support on $[0, 1]$, where $\tau_0 = \tau_g + \vartheta$. Further insights may be gained using the tractable Gaussian learning from Assumption 3.5. Indeed, in this case $f_{\tilde{\tau}}(\tau|\tau_0; w)$ is a Gaussian pdf $\phi(\tau; \mu, \sigma^2)$ with mean $\mu = \xi\tau_0 + (1 - \xi)\hat{\tau}$ and variance $\sigma^2 = \sigma^{\star 2}$. We can thus express agents' responses to tax reforms in terms of changes in the true tax rate τ_0, changes in the prior mean $\hat{\tau}$ and induced changes in attention ξ associated with changes in the precision of the signal $\sigma^\star$. To that end, we use a first-order approximation of the objective at the mean μ and exploit the following Lemma.

Lemma 1. *Let $\psi(x)$ be a differentiable real-valued function, $\psi_a(x) = \psi(a) + (x - a)\psi'(a)$ its first-order Taylor approximation evaluated at a and $\phi(x; \mu, \sigma^2)$ the pdf of the gaussian distribution with mean μ and variance σ^2. Then,*

$$\int_{\mathbb{R}} \psi_\mu(x)\phi(x; \mu, \sigma^2)dx = \psi(\mu) \tag{A.10}$$

$$\int_{\mathbb{R}} \psi_\mu(x)\frac{\partial\phi(x; \mu, \sigma^2)}{\partial\mu}dx = \psi'(\mu) \tag{A.11}$$

$$\int_{\mathbb{R}} \psi_\mu(x)\frac{\partial\phi(x; \mu, \sigma^2)}{\partial\sigma}dx = 0 \tag{A.12}$$

Proof. Equation (A.10) directly follows from $\int_{\mathbb{R}}(x - \mu)\phi(x; \mu, \sigma^2) = 0$ by definition of the mean. To prove equation (A.11), realize that $\int_{\mathbb{R}} \frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu}dx = 0$ and $\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu} = \frac{x-\mu}{\sigma^2}\phi(x; \mu, \sigma^2)$ so that $\int_{\mathbb{R}}(\psi(\mu) + (x - \mu)\psi'(a))\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu}dx = \frac{\psi'(\mu)}{\sigma^2}\int_{\mathbb{R}}(x - \mu)^2\phi(x; \mu, \sigma^2) = \psi'(\mu)$. Equation (A.12 follows from the fact that $\int_{\mathbb{R}} \frac{\partial\phi(x;\mu,\sigma^2)}{\partial\sigma}dx = 0$ such that the integral of a constant is nil and that $\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\sigma}$ is symetric such that the integral of x also nil by a symetry argument. □

Rewriting equation (A.7) as

$$E_\vartheta\Big[\int\Big\{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star+y^\star\Big]\phi(\tau;\mu,\sigma^2)d\tau \qquad \text{(A.13)}$$
$$+\int\Big[\frac{G(\mathcal{V})}{p}+\tau_0 y^\star\Big]\Big(\frac{d\phi(\tau;\mu,\sigma^2)}{d\mu}\frac{d\mu}{d\tau_g}+\frac{d\phi(\tau;\mu,\sigma^2)}{d\sigma}\frac{d\sigma}{d\tau_g}\Big)d\tau\Big\}f_w(w)dw\Big]=0$$

allows us to apply Lemma 1 to obtain with $\mu=\xi\tau_0+(1-\xi)\hat{\tau}$

$$E_\vartheta\Big[\int\Big\{\Big[-\frac{G'(\mathcal{V})}{p}y^\star+y^\star\Big]\Big|_{\tilde{\tau}=\mu} \qquad \text{(A.14)}$$
$$+\Big[\Big(\frac{G'(\mathcal{V})}{p}(\tilde{\tau}-\tau_0)+\tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\frac{d\mu}{d\tau_g}\Big]\Big|_{\tilde{\tau}=\mu}\Big\}f_w(w)dw\Big]=0$$

since taking $\psi(\tau)=\Big[\frac{G(\mathcal{V})}{p}+\tau_0 y^\star\Big](\tau)$ implies $\psi'(\mu)=\Big[\Big(\frac{G'(\mathcal{V})}{p}(\tilde{\tau}-\tau_0)+\tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\Big](\mu)$ by the modified envelope condition. Now, in equilibrium we have by definition that $\hat{\tau}=\tau_g+b$ meaning $\mu=\tau_g+\xi\vartheta+(1-\xi)b$ and $\mu-\tau_0=(1-\xi)(b-\vartheta)$. Hence, in equilibrium,

$$E_\vartheta\Big[\int\Big\{\Big[-\frac{G'(\mathcal{V})}{p}y^\star+y^\star\Big]\Big|_{\tilde{\tau}=\tau_g+\xi\vartheta+(1-\xi)b} \qquad \text{(A.15)}$$
$$+\Big[\Big(\frac{G'(\mathcal{V})}{p}(1-\xi)(b-\vartheta)+\tau_g+\vartheta\Big)\frac{dy^\star}{d\tilde{\tau}}\frac{d\mu}{d\tau_g}\Big]\Big|_{\tilde{\tau}=\tau_g+\xi\vartheta+(1-\xi)b}\Big\}f_w(w)dw\Big]=0$$

Last, we characterize taxpayers' average response to tax reforms $\frac{d\mu}{d\tau_g}$ as computed by the government and by the social planner. The government takes agent's priors and thus attention strategies as given, hence $\frac{d\mu}{d\tau_g}=\xi\frac{d\tau_0}{d\tau_g}=\xi$ which yields equation (18). In contrast, the social planner internalizes the equilibrium condition that priors and thus attention strategies adjust to the government policy such that $\frac{d\mu}{d\tau_g}=\xi\frac{d\tau_0}{d\tau_g}+(1-\xi)\frac{d\hat{\tau}}{d\tau_g}+\frac{d\xi}{d\tau_g}(\tau_0-\hat{\tau})=1+\frac{d\xi}{d\tau_g}(\vartheta-b)$ in equilibrium. This yields equation (23).

Equations (19) and (23) follow from a direct application of Lemma 1 to equation A.9 with again $\mu=\tau_g+\xi\vartheta+(1-\xi)b$:

$$E_\vartheta\Big[\int\frac{G'(\mathcal{V})}{p}\Big|_{\tilde{\tau}=\mu}f(w)dw\Big]=1 \qquad \text{(A.16)}$$

A.5 Sufficient statistics formulas in tractable Gaussian case

Taking a small noise approximation, the characterization of equilibrium positive tax rate τ_g^{eq} and normative tax rate $\tau_g^\star$ in this tractable Gaussian model write

$$\int\Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star+\Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^{\text{eq}}\Big)\frac{dy^\star}{d\tilde{\tau}}\xi\Big]\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b}f_w(w)dw=0$$
$$\int\Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star+\Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^\star\Big)\frac{dy^\star}{d\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}b\Big)\Big]\Big|_{\tilde{\tau}=\tau_g^\star+(1-\xi)b}f_w(w)dw=0$$

Assuming preferences are iso-elastic e.g. $U(c,y;w) = c - \frac{(y/w)^{1+\varepsilon}}{1+\varepsilon}$, the elasticity of labor supply responses with respect to the perceived marginal net-of-tax rate e is constant

$$\forall \tilde{\tau}, w, \quad e \equiv \frac{1-\tilde{\tau}}{y^\star}\frac{dy^\star}{d(1-\tilde{\tau})} = \frac{1}{\varepsilon} \iff \frac{dy^\star}{d\tilde{\tau}} = -e\frac{y^\star}{1-\tilde{\tau}} \tag{A.17}$$

Plugging in e we get

$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star - \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b + \tau_g^{\text{eq}}\Big)e\frac{y^\star}{1-\tilde{\tau}}\xi\Big]\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b} f_w(w)dw = 0$$
$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star - \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b + \tau_g^\star\Big)e\frac{y^\star}{1-\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}b\Big)\Big]\Big|_{\tilde{\tau}=\tau_g^\star+(1-\xi)b} f_w(w)dw = 0$$

To further simplify these formulas we now make a small perception bias approximation $b << 1$. This allows us to use the approximation $\frac{1}{1-\tau_g-(1-\xi)b} \approx \frac{1}{1-\tau_g}$ and to assume $\frac{d\xi}{d\tau_g}b << 1$ to simplify some terms[I]. Defining social marginal welfare weights $g(w) \equiv \frac{G'(\mathcal{V})}{p}$ and the mean operator $\bar{x} = \int x(w)f(w)dw$ we get

$$\Big\{\overline{(1-g)y^\star} - \frac{\tau_g^{\text{eq}}}{1-\tau_g^{\text{eq}}}\overline{y^\star\xi}e - \frac{b}{1-\tau_g^{\text{eq}}}\overline{g(1-\xi)y^\star\xi e}\Big\}\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b} = 0$$
$$\Big\{\overline{(1-g)y^\star} - \frac{\tau_g^\star}{1-\tau_g^\star}\overline{y^\star}e - \frac{b}{1-\tau_g^\star}\overline{g(1-\xi)y^\star e}\Big\}\Big|_{\tilde{\tau}=\tau_g^\star+(1-\xi)b} = 0$$

which simplify to the compact sufficient statistics formulas

$$\tau_g^{\text{eq}} = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\,\frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{A.18}$$
$$\tau_g^\star = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\,\frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{A.19}$$

where all endogenous quantities on the right hand-side of the equations are evaluated at respectively $\tilde{\tau} = \tau_g^{\text{eq}} + (1-\xi)b$ and $\tilde{\tau} = \tau_g^\star + (1-\xi)b$. In other words formulas are expressed in terms of sufficient statistics evaluated at the optimum.

A.6 Taxation bias in tractable Gaussian case

A difficulty in comparing τ_g^{eq} and $\tau_g^\star$ is that some right-hand side quantities are endogenous to the tax rate and thus evaluated at different tax rates. To overcome this difficulty, we use a small taxation bias approximation $\tau_g^{\text{eq}} \approx \tau_g^\star = t$ such that quantities can be evaluated to

[I] In our simulations we do check that $\frac{d\xi}{d\tau_g}$ does not take large values (it takes values between 0.2 and 1 in equilibrium) as a way to confirm the validity of this approximation.

a first-order approximation at the same tax rate. Furthermore, we assume that corrective motives associated to the presence of a perception bias b are evaluated at tax rate t such that we can finally directly compare

$$\tau_g^{\text{eq}} = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\ \frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{A.20}$$

$$\tau_g^\star = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\ \frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{A.21}$$

The first terms on the right-hand side corresponds to the standard optimal tax formula (e.g. Piketty and Saez (2013)) whereas the second are corrective terms associated to the existence of a perception bias b. For small perception biases, these corrective terms are second-order and go in the same direction for both positive and normative tax rates. They are thus not driving the difference between the two and we disregard them to derive the following closed form formula for the taxation bias

$$\begin{aligned}
\tau_g^{\text{eq}} - \tau_g^\star &= \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} && \text{(A.22)}\\
&= \frac{\overline{(1-g)y^\star}\left(\overline{(1-g)y^\star} + \overline{y^\star}\ e\right) - \overline{(1-g)y^\star}\left(\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e\right)}{\left(\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e\right)\left(\overline{(1-g)y^\star} + \overline{y^\star}\ e\right)} && \text{(A.23)}\\
&= \frac{e\ \tau_g^{\text{eq}}\ \tau_g^\star}{\overline{(1-g)y^\star}}\left(\overline{y^\star} - \overline{y^\star\xi}\right) && \text{(A.24)}
\end{aligned}$$

A.7 Utility decomposition

Let $\mathcal{V}_\kappa(\tau_g^\star(\kappa), w)$ be the expected utility of taxpayer w at the positive equilibrium when the information cost is κ and the optimal target tax rate of the government is $\tau_g^\star(\kappa)$. Then, from Assumption 3.3,

$$\begin{aligned}
\mathcal{V}_0(\tau_g^\star(0), w) &= E_{\tau_0|\tau_g^\star(0)}\Big[R_0 + (1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)\Big]\\
\mathcal{V}_\kappa(\tau_g^\star(\kappa), w) &= E_{\tau_0|\tau_g^\star(\kappa)}\Big[\int \Big(R_0 + (1-\tau_0)y^\star(\tau; w) - v(y^\star(\tau; w); w)\Big) f_{\tilde{\tau}}(\tau|\tau_0, w)d\tau\Big] - \kappa\mathcal{I}(\sigma^\star(\hat{q}(\tau), \kappa, w))
\end{aligned}$$

Using straightforward algebra,

$$\mathcal{V}_\kappa(\tau_g^\star(\kappa), w) - \mathcal{V}_0(\tau_g^\star(0), w) = \Delta^{\mathcal{V}}R + \Delta^{\mathcal{V}}\tau + \Delta^{\mathcal{V}}b + \Delta^{\mathcal{V}}\text{uncertainty} + \Delta^{\mathcal{V}}\text{info cost} \tag{A.25}$$

where

$$\Delta^{\mathcal{V}}R \equiv E_{\tau_0|\tau_g^\star(\kappa)}[R_0] - E_{\tau_0|\tau_g^\star(0)}[R_0]$$

is the change in the average demogrant,

$$\Delta^{\mathcal{V}}\tau \equiv E_{\tau_0|\tau_g^\star(\kappa)}[(1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)] - E_{\tau_0|\tau_g^\star(0)}[(1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)]$$

is the change in the expected utility due to the change in the tax target $\tau_g^\star$,

$$\begin{aligned}\Delta^{\mathcal{V}} b \;\; \equiv \;\; & E_{\tau_0|\tau_g^\star(\kappa)}\big[(1-\tau_0)\big(y^\star(\tau_0+(1-\xi)b;w) - y^\star(\tau_0;w)\big) \\ & -\big(v(y^\star(\tau_0+(1-\xi)b;w);w) - v(y^\star(\tau_0;w);w)\big)\big]\end{aligned}$$

is the change in the expected utility due to the bias b,

$$\begin{aligned}\Delta^{\mathcal{V}}\text{uncertainty} \;\; \equiv \;\; & E_{\tau_0|\tau_g^\star(\kappa)}\Big[\int (1-\tau_0)\big(y^\star(\tau;w) - v(y^\star(\tau;w);w\big) \\ & \phi\big(\tau;\xi\tau_0+(1-\xi)(\tau_g^\star(\kappa)+b),(\xi\sigma^\star)^2\big)d\tau \\ & -\big((1-\tau_0)\big(y^\star(\tau_0+(1-\xi)b;w) - v(y^\star(\tau_0+(1-\xi)b;w);w)\big)\Big]\end{aligned}$$

is the change in the expected utility due to noisy information and $\Delta^{\mathcal{V}}\text{info cost} = -\kappa\mathcal{I}(\sigma^\star(\hat{q}(\tau),\kappa,w))$.