Optimal Taxation with Rational Inattention

George-Marios Angeletos   Matias Bayas-Erazo

January 20, 2026

Abstract

We study the implications of rationally inattentive behavior for the design of optimal taxes. Our main finding is an irrelevance result: when inattention satisfies two key properties, "invariance" and "state separability", optimal taxes satisfy exactly the same type of sufficient-statistics formula as that found in classical public finance. Violations of state-separability generate interdependence of the optimal taxes across states but they do not change the spirit of optimal tax formulas. Departures from invariance, on the other hand, open the door to cognitive externalities and call for taxes that help simplify the price system. We also show that inattention does not necessarily make agents less responsive and may thus justify *lower* taxes.

1 Introduction

How does inattention affect the design of optimal taxes? Because inattention distorts market outcomes relative to the first best, one may argue that optimal taxes should aim at undoing or correcting this distortion. On the other hand, one could also argue that, in a second-best world, a benevolent planner may want to balance the effects of inattention with the effects of distortionary taxation, or perhaps even exploit people's inattention to prices and taxes so as to extract more revenue with less tax distortion.

Both of these ideas, versions of which have appeared in behavioral public finance (Goldin, 2015; Gerritsen, 2016; Allcott et al., 2018; Farhi and Gabaix, 2020), seem to presume that inattention is a distortion akin to an externality. But this makes little sense if inattention is the result of rational behavior in the presence of information costs or cognitive constraints, as in the literature pioneered by Sims (2003). In this paper, we therefore ask: how does *rational* inattention affect the design of optimal taxes?

We address this question by augmenting the classic, multi-good, Ramsey taxation problem (Diamond, 1975) with a flexible form of rational inattention. Consumers make their choices on the basis of an imperfect signal of the relevant fundamentals, prices, and taxes. This signal—and the noise therein—is a representation of costly information acquisition, costly cognitive effort, or costly optimization. The consumer chooses the precision of this signal to balance the benefits of smaller mistakes in consumption against the costs of more information or more cognitive effort. Finally, these costs can have a very flexible form, nesting not only mutual information costs (Sims, 2010; Maćkowiak et al., 2023) but also a variety of other specifications proposed in an emerging decision-theoretic and experimental literature.

Our main finding (Theorem 1) is an irrelevance result: when inattention satisfies two key properties, "invariance" à la Angeletos and Sastry (2025) and "state separability" à la Flynn and Sastry (2023), the optimal taxes satisfy exactly the same type of sufficient-statistics formula as that found in classical public finance. In simpler words, the analyst or policy maker can totally ignore the presence of inattention and proceed as usual: measure the price elasticities of the aggregate demand for the different commodities, plug these elasticities in familiar optimal taxation formulas, and compute the implied optimal taxes, as if people were fully attentive.

While this irrelevance result depends on strong assumptions, in our view it represents an important benchmark for three reasons. First, it clarifies that the argument for corrective taxation found in some of the behavioral literature hinges on a paternalistic perspective, whereby the planner does not internalize people's attention costs and treats inattention as an inherent "bad".

Second, it verifies the validity of the second argument mentioned above, namely that in favor of exploiting people's inattention to raise more tax revenue, but qualifies it in the following way: inattention may *increase* some demand elasticities while decreasing others. In fact, in the simplest formulation of costly control, where agents face difficulty implementing their desired consumption choices precisely, inattention systematically *increases* effective demand elasticities. This contradicts the conventional view that inattention makes consumers less responsive and thus calls for higher taxes; instead, it suggests that optimal taxes may be *lower* than in the fully attentive case. Regardless of the direction, all that matters are the measured demand elasticities, not the inattention or preferences behind them.

Last but not least, this benchmark helps organize the *novel* considerations for optimal taxes that rational inattention can introduce when, and *only when*, the aforementioned two assumptions are violated.

Violations of state-separability generate interdependence of optimal taxes across states. Formally, the optimal taxes in a given state depend not only on the demand elasticities and the shadow value of tax revenue in that state (as in standard PF), but also on the values of these objects in all other states. The spirit of optimal tax formulas, however, remains unchanged. In fact, it is analogous to the formula one would obtain in a fully attentive economy with preferences that are not separable across states. At the same time, this interdependence does introduce a significant empirical challenge: measuring cross-state elasticities of demand. This requires the right kind of tax experiment, as standard approaches that measure elasticities using variation across states would lead to incorrect estimates of the relevant objects. Furthermore, when attention costs create linkages across states, the planner may optimally use state-dependent taxes to influence attention allocation. We show that the direction of this adjustment depends on the elasticity of demand.

Violations of invariance introduce truly novel considerations for optimal taxation. They open the door to cognitive externalities (Angeletos and Sastry, 2025) and help accommodate the idea that markets may be excessively complex, even when firms are perfectly competitive.

In this case, the optimal taxes may aim at "simplifying prices" or otherwise helping consumers minimize their mistakes. For example, when higher price volatility increases attention costs (a positive cognitive externality), the planner optimally uses taxes to compress the price distribution, reducing the cognitive burden on agents. The optimal tax system now balances two goals: easing the externality (Pigou) and minimizing the distortion from raising revenue (Ramsey). Finally, departures from invariance help accommodate the role of salience, and in particular the idea that it may make a crucial difference for efficiency and welfare whether prices are quoted in pre-tax terms (as in the US) or post-tax terms (as in Europe).

Related literature

Our paper connects to the rational inattention literature and to behavioral public finance.

First, we build on the rational inattention framework pioneered by Sims (2003) and surveyed in Maćkowiak et al. (2023). Caplin and Dean (2015) and Caplin et al. (2022) provide foundations for information cost functions and characterize the behavioral implications of posterior-separable costs. More recent work studies the efficiency properties of economies with inattentive agents. Angeletos and Sastry (2025) extend the welfare theorems to economies with rationally inattentive agents, showing that invariance is sufficient for efficiency and that violations of invariance can generate cognitive externalities. Hébert and La'O (2023) obtain related efficiency characterizations in large games with endogenous information acquisition. We build on these insights, but our focus on optimal taxation, including the sufficient-statistics formulas and the role of cognitive externalities, is new.

Second, we contribute to the behavioral public finance literature. A growing literature studies how tax salience and misperceptions affect incidence and optimal policy (Chetty et al., 2009; Finkelstein, 2009; Goldin, 2015; Allcott et al., 2018) and develops optimal tax formulas with general behavioral wedges (Farhi and Gabaix, 2020) or under alternative welfare criteria (Gerritsen, 2016). Unlike these papers, we treat inattention as an endogenous choice and use a flexible framework that can speak to a range of behavioral phenomena. Moreover, we incorporate attention costs directly into welfare, so there is no separate corrective motive under the conditions we study. Closer to us, Boccanfuso and Ferey (2023) study how inattention to taxes generates a time-inconsistency problem for tax policy.

2 A SIMPLE MODEL OF RATIONAL INATTENTION

This section introduces a simple framework to study optimal commodity taxation when agents are rationally inattentive. Our goal is to illustrate the key features of rational inattention and their implications for optimal taxation in a transparent way. We begin by describing the basic economic environment and a general formulation of the optimization problem faced by inattentive agents. We then provide two concrete examples that capture different types of attention costs, before turning to the government's optimal policy problem.

2.1 BASIC ENVIRONMENT

Consider a quasilinear economy with two goods: a taxable good x and an untaxable numeraire good y (with price normalized to one). There is a stochastic state of nature $\theta \in \Theta$ with prior distribution $\pi_\Theta \in \Delta(\Theta)$. The state θ determines the producer price $p(\theta)$ of good x. The government sets a tax rule $\tau : \Theta \to \mathbb{R}$, so that the after-tax price of good x in state θ is $q(\theta) \equiv p(\theta) + \tau(\theta)$.

From the perspective of agents, the after-tax price q is random and difficult to observe or process. Agents reason about both the after-tax price of good x and the exogenous payoff-relevant fundamental θ. We refer to this object as the *cognition state* $z \equiv (\theta, q)$. This captures the idea that agents care about both the underlying economic fundamentals and prices when making consumption decisions. Our results extend to alternative specifications of the cognition state, such as $z = (\theta, p, \tau)$, but we focus on this case in the main text for simplicity.

Agents have prior beliefs π over the cognition state z. We say that the prior is *consistent* if $\pi(z) = \pi_\Theta(\theta)\mathbb{1}_{q(\theta)}(q)$, where $\mathbb{1}_{q(\theta)}(q)$ is an indicator function that equals one when $q = q(\theta)$ and zero otherwise. In words, the prior is consistent when it reflects the relationship between fundamentals and prices. This consistency requirement is standard in rational expectations equilibria and ensures that agents' beliefs about the environment are in line with the actual policy chosen by the government. We require consistency of the prior throughout the rest of the paper. Finally, we let $Z_\pi \equiv \{z : \pi(z) > 0\}$ denote the support of the prior.

2.2 Individual problem

We model inattention as the choice of a state-dependent stochastic choice rule (Stahl, 1990; Fudenberg et al., 2015; Morris and Yang, 2022). This rule describes how an agent's consumption varies with the cognition state. Specifically, agents choose $F = \{F(x \mid z)\}_{z \in Z_\pi}$, where $F(x \mid z)$ describes the cumulative distribution of consumption of good x in cognition state z. We assume that the consumption of good y adjusts to ensure that the agent's budget constraint holds.

This formulation captures the idea that inattentive agents cannot perfectly condition their behavior on the state of the world they are trying to understand. Instead, they optimally choose the degree of randomness in their consumption decisions, trading off the benefits of more precise control against the costs of achieving such precision. As discussed by Denti (2023), this class of models nests all models of costly information acquisition that have a posterior separable representation.

The individual problem is to choose F in order to maximize:

$$\sum_{z \in Z_\pi} \int_{\mathcal{X}} \{u(x, \theta) + w - qx\} \, dF(x \mid z) \pi(z) - C(F, \pi), \tag{1}$$

where u is a Bernoulli utility function for good x, w is income, and C is a cost function that captures the costs of cognitive effort, information acquisition, or precise control. We assume that, conditional on each state θ, the utility function $u(\cdot, \theta)$ is strictly increasing and concave in x and that the cost function satisfies appropriate regularity conditions to ensure the problem is well-defined. Notice that the cost function may depend on the prior π, reflecting the idea that the cognitive cost varies with what the agent already knows. This dependence on the prior π is what makes taxation with inattentive agents non-trivial: the government's choices affect not only prices but also how agents allocate their attention.

This formulation is deliberately flexible and allows us to capture different models of inattention without committing to a specific microfoundation. In particular, it is consistent with models of costly cognition and costly control. In the former interpretation, agents acquire signals about the state and then choose consumption based on these signals. In the latter, agents have difficulty implementing their desired consumption plans and must optimally choose the precision of their control. While this flexibility comes at the cost of appearing somewhat reduced-form, it allows us to study optimal commodity taxation across a broad class of decision frictions. To make the framework more concrete, we now present two examples that highlight different types of attention costs and their implications.

2.3 Two illustrative examples

The following examples serve two purposes: they provide more familiar formulations of the inattention problem, and they show how the structure of attention costs can lead to different implications for optimal taxation. The key distinction we emphasize is between costs that satisfy certain invariance and separability properties and those that do not. We will return to these examples throughout the paper to illustrate our main results.

Example 1: Costly cognitive effort in a 2×2 economy

Our first example builds on the rational inattention literature pioneered by Sims (2003). Suppose there are two possible states, $\theta \in \{H, L\}$, with $\pi_\Theta(H) = \pi_H$ and $\pi_\Theta(L) = 1 - \pi_H$. Agents are initially confused about the underlying state and receive a signal $\omega \in \{H, L\}$. They choose the signal structure $\phi = \{\phi(\omega \mid z)\}_{z \in Z_\pi}$ that determines the probability of receiving each signal conditional on the cognition state. The signal structure in this 2×2 economy can be described with a matrix whose entries represent the probabilities of receiving each signal in each state, as shown in Figure 1.

Figure 1: Signal structure in the 2×2 economy

	$\omega = L$	$\omega = H$
$z = (L, q(L))$	$1 - \epsilon_L$	ϵ_L
$z = (H, q(H))$	ϵ_H	$1 - \epsilon_H$

In state $z = (L, q(L))$, the agent receives a signal $\omega = L$ with probability $1 - \epsilon_L$ and the signal $\omega = H$ with probability ϵ_L. In state $z = (H, q(H))$, the agent receives a signal $\omega = H$ with probability $1 - \epsilon_H$ and the signal $\omega = L$ with probability ϵ_H. Thus, ϵ_L and ϵ_H capture the probability of making a mistake in each state.

After observing their signal, agents choose consumption according to a deterministic function $x(\omega)$ that maps signal realizations to consumption levels. Their objective is to maximize:

$$\sum_{z \in Z_\pi} \sum_{\omega \in \{H,L\}} \{u(x(\omega)) + w - qx(\omega)\} \, \phi(\omega \mid z)\pi(z) - C(\phi, \pi). \tag{2}$$

Following the Sims (2003) tradition, we assume that cognitive costs C depend on the mutual information between the signal ω and the cognition state z, denoted by $\mathcal{I}(\omega, z)$. Intuitively, mutual information measures the reduction in uncertainty about z upon observing ω. The more precise is the information contained in the signal, the higher is the mutual information between ω and z. The key property of mutual information costs is that they are *invariant* to relabelings of the state space. This invariance property, which we define formally in Section 3, will prove crucial for our main irrelevance result.

By the revelation principle, this two-step formulation of the individual problem can be transformed into the choice of a state-dependent stochastic choice rule as in (1), where agents directly choose a distribution of actions in each state (Matějka and McKay, 2015; Denti, 2023).

EXAMPLE 2: COSTLY CONTROL

Our second example illustrates a different type of decision friction based on the idea of *costly control*. Here, agents are no longer confused about the state but instead have difficulty mapping states to optimal actions. Specifically, we assume that consumption of good x in state z equals $X(z)e^{\epsilon}$, where $X(z)$ denotes a target consumption level in state z and $\epsilon \sim \mathcal{N}(0, \sigma_{\epsilon}^2(z))$ is a tremble. The variance of the tremble $\sigma_{\epsilon}^2(z)$ is optimally chosen by the agents and determines how precisely agents can control their consumption in each state. This captures the idea that it is hard for agents to figure out the right amount of consumption in each state.

Conditional on state z, agents choose both their target consumption level $X(z)$ and the variance of the tremble $\sigma_{\epsilon}(z)$ to maximize:

$$\mathbb{E}\left[u(X(z)e^{\epsilon}) - q \cdot X(z)e^{\epsilon} - C(\sigma_{\epsilon}(z))\right], \tag{3}$$

where $C(\sigma_{\epsilon}(z))$ captures the cost of precise control in state z. Lower values of $\sigma_{\epsilon}(z)$ correspond to more precise control and higher costs.

The key property of this specification is that attention costs are separable across states. This means that control costs in state z are independent of what happens in any other state z'. For this reason, we can study the individual problem state-by-state. This separability stands in contrast to mutual information costs, which typically create linkages across states. This family of control costs, which we define formally in Section 3, was introduced in a recent paper by Flynn and Sastry (2023), who show that they can be used to model several kinds of decision frictions. As we will see, the distinction between separable and non-separable attention costs also plays an important role in our irrelevance result.

2.4 The government's planning problem

Having described the economy, we now turn to the government's optimal policy problem. To formulate the problem, we need to introduce some additional notation. Let $\bar{x}(\theta; \boldsymbol{Q})$ denote aggregate demand for good x in state θ:

$$\bar{x}(\theta; \boldsymbol{Q}) \equiv \int_{\mathcal{X}} x \, dF(x \mid z_\theta; \boldsymbol{Q}), \tag{4}$$

where $z_\theta \equiv (\theta, q(\theta))$ for all $\theta \in \Theta$. Here, we make explicit that the distribution F may depend on the entire vector of after-tax prices $\boldsymbol{Q} \equiv \{q(\theta')\}_{\theta' \in \Theta}$, not just the price in state θ. Additionally, let $v(\boldsymbol{Q}, w, \pi)$ denote the consumer's indirect utility function. It depends on the price vector $\boldsymbol{Q}$ and income w, as in standard models, as well as the prior distribution π over cognition states.

The government chooses a tax rule $\tau : \Theta \to \mathbb{R}_+$ to maximize social welfare subject to two constraints:

$$\begin{aligned} \max_{\tau} \quad & v(\boldsymbol{Q}, w, \pi) + \lambda \sum_{\theta \in \Theta} \pi_\Theta(\theta) \tau(\theta) \bar{x}(\theta; \boldsymbol{Q}) \qquad (5) \\ \text{s.t.} \quad & \pi(z) = \pi_\Theta(\theta) \mathbb{1}_{q(\theta)}(q), \quad \forall z = (\theta, q) \in Z_\pi, \\ & F \in \arg\max_{\tilde{F}} \sum_{z \in Z_\pi} \int_{\mathcal{X}} \{u(x, \theta) + w - qx\} \, d\tilde{F}(x \mid z) \pi(z) - C(\tilde{F}, \pi). \end{aligned}$$

The first constraint ensures that agents' prior beliefs are consistent with the tax rule. The second is the implementability condition: it requires that the stochastic choice rule F be optimal for the consumer given the price system and consistent prior. The parameter $\lambda > 1$ represents the shadow value of public funds. For simplicity, we treat λ as a parameter that governs the spending needs of the government. As usual, the Ramsey planner tries to raise revenue while minimizing the distortionary effects of taxation on consumer welfare. When λ increases, the planner places a higher weight on revenue relative to consumer welfare.

This planning problem differs from the standard Ramsey problem with fully attentive agents in two key ways. These differences highlight the challenges that rational inattention introduces for optimal tax design.

First, the aggregate demand in state θ, $\bar{x}(\theta; \boldsymbol{Q})$, may depend on the entire vector of after-tax prices $\boldsymbol{Q}$, not just $q(\theta)$. When agents are fully attentive and have preferences that are separable across states, this is not the case: aggregate demand in state θ depends only on the after-tax price in the given state $q(\theta)$. Under rational inattention, however, agents may use information about prices in some states to infer prices in others.

To see this more clearly, consider an agent who has difficulty distinguishing between different states, as in Example 1 above. In such a world, the agent will receive a signal and use this to update beliefs about the after-tax price. When the agent cannot perfectly distinguish between states, their posterior belief about prices will depend on the price of good x in both state H and state L. Hence, changing the after-tax price in, say, state H will affect aggregate demand both in state H and in state L. This leads to a form of *interdependence* that is typically absent with fully attentive agents.

Second, taxes affect consumer welfare not only through their impact on after-tax prices but also through their effect on the prior distribution π. Recall that this prior is generated by combining the exogenous prior over fundamentals π_Θ with the "true" after-tax prices, which are pinned down by the tax policy chosen by the government. In particular, when the prior is consistent, the tax rule τ determines which cognition states $z = (\theta, q)$ are possible, thereby shaping the agent's cognitive environment. Changes in the prior distribution π may thus alter the attention costs $C(F, \pi)$ faced by agents.

For example, a tax policy that creates very similar after-tax prices across states with different fundamentals may reduce the cognitive burden on agents whenever attention costs depend on the variability of after-tax prices. Conversely, a policy that creates more dispersed prices may increase it. This opens the door to a novel channel through which tax policy affects welfare–one that operates through the cognitive burden imposed on agents, rather than solely through traditional substitution and income effects.

Together, these two differences suggest that the planning problem under rational inattention is inherently high-dimensional. The effect of taxes on social welfare is hard to compute because they enter indirect utility via the prior and because they affect aggregate demand across all states. However, as we show next, our notions of invariance and state separability allow us to side-step these considerations entirely. Specifically, invariance mutes the dependence of indirect utility on the endogenous prior π, while state separability eliminates the cross-state dependence of aggregate demand. Together, they deliver a simple characterization of optimal commodity taxes with inattentive agents.

3 THE IRRELEVANCE RESULT

We now present our main irrelevance result: under two restrictions on attention costs, *invariance* and *state separability*, rational inattention is irrelevant for optimal commodity taxation. Specifically, the optimal taxes satisfy exactly the same type of sufficient-statistics formulas as in classical public finance. Inattention matters exclusively through its impact on effective elasticities.

We first formalize these two properties and explain their economic interpretation. We then state our irrelevance result for the simple two-good economy described in Section 2 and provide some intuition. Finally, we illustrate the theorem's implications through our costly control example and extend the result to a general economy with multiple goods and heterogeneous agents.

3.1 DEFINING STATE SEPARABILITY AND INVARIANCE

We begin by formalizing the two properties of attention costs that drive our irrelevance result. These definitions are adapted from Flynn and Sastry (2023) and Angeletos and Sastry (2025) to our setting. While they are somewhat technical, the underlying economic interpretations are straightforward.

3.1.1 STATE SEPARABILITY

Our first restriction captures the idea that attention costs exhibit a form of separability across different states of the world.

Definition 1. A cost functional C is *state separable* if there exists a strictly convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$ and a weighting function $\mu : \Theta \times \mathbb{R}^N_+ \to \mathbb{R}_{++}$ such that for any F with density f:

$$C(F, \pi) = \sum_{z \in Z_\pi} \mu(z)\pi(z) \int_{\mathcal{X}} \phi(f(x \mid z))dx. \tag{6}$$

If F does not admit a density, we replace $\int_{\mathcal{X}} \phi(f(x \mid z))dx$ with $\sum_{x \in \mathcal{X}} \phi(\mathbb{P}(x \mid z))$.

Intuitively, state separability requires the cost of controlling mistakes in state z to be independent of what happens in state z'. For example, consider an agent deciding how much to consume when facing different after-tax prices across states. State separability captures the idea that the cognitive difficulty of controlling consumption in one state (say, during a recession) is independent of how precise consumption is in another state (say, during a boom). The costs of controlling mistakes in different states may depend on the identity of those states through the weighting function μ but not on the choices made in other states. Convexity of the function ϕ ensures that it is costly to consume with a high degree of precision in any state.

This property reflects a "local" view of inattention, where cognitive resources can be allocated independently across different economic environments. It's as if the agent maintains separate mental accounts for different situations, effectively facing separate attention allocation problems across different states. This implies that the agent's problem can be solved state-by-state, without the need to consider how choices in one state affect choices in another. Although state-separability may seem restrictive, it can capture a wide range of decision frictions, including ex-post optimization errors, ex-ante planning frictions, and endogenous consideration sets (cf. Flynn and Sastry, 2023). This flexibility is achieved by changing the weighting function μ and/or the kernel ϕ.

3.1.2 Invariance

Our second key property requires that attention costs be invariant to certain transformations of the environment. To define this precisely, we must first introduce transformations of random plans of actions and an appropriate notion of sufficiency. These concepts help formalize the idea that the same underlying decision problem can be described in multiple ways, and invariance requires that attention costs remain unchanged across equivalent descriptions.

Definition 2 (Transformation)**.** Consider two random plans of action (F, π) and $(\tilde{F}, \tilde{\pi})$ and a mapping $g : \Theta \times \mathbb{R}_+^N \to \Theta \times \mathbb{R}_+^N$. We say that $(\tilde{F}, \tilde{\pi})$ is the *transformation* of (F, π) under g if:

$$\tilde{\pi}(z) = \sum_{z' \in Z_\pi} \pi(z') \mathbb{1}_z\{g(z')\}, \quad \forall z \in Z_\pi, \tag{7}$$

$$\tilde{f}(x \mid z) = \frac{\sum_{z' \in Z_\pi} f(x \mid z') \pi(z') \mathbb{1}_z\{g(z')\}}{\tilde{\pi}(z)}, \quad \forall x, z. \tag{8}$$

Intuitively, a transformation relabels or regroups states of the world via the mapping g. This can be thought of as a change of variables that might, for example, rename states or combine finely detailed states into coarser categories. The first part of the definition describes the distribution of this new random variable. The second part specifies how the conditional distribution of actions should be constructed from the original plan (F, π) to preserve its informational content. Together, these two conditions ensure that a transformation changes how the environment is described while ensuring that no information is lost beyond what is captured by the mapping g.

Definition 3 (Sufficiency)**.** Consider two random plans of action (F, π) and $(\tilde{F}, \tilde{\pi})$ such that $(\tilde{F}, \tilde{\pi})$ is a transformation of (F, π) under some mapping g. We say that $\tilde{\pi}$ is *sufficient* for π with respect to F if $F(x \mid z) = \tilde{F}(x \mid g(z))$ for all x and all $z \in Z_{\pi}$.

In words, sufficiency means that the transformed plan generates the same conditional distributions of actions as the original plan, but possibly under a different labeling or aggregation of states. The equality $F(x \mid z) = \tilde{F}(x \mid g(z))$, implies that the transformed state $g(z)$ is "sufficient" for z in the sense that it carries all the information from z that is relevant for predicting the agent's actions. Therefore, while the transformation g might change how states are described (e.g., by relabeling or aggregating them), a sufficient transformation ensures that this does not alter the statistical relationship between states and actions.

With these definitions in place, we are now ready to define our notion of invariance with respect to specific classes of transformations:

Definition 4 (Invariance)**.** Fix a set $G \subseteq \{g : \Theta \times \mathbb{R}_+^N \to \Theta \times \mathbb{R}_+^N\}$. Consider any functioon $g \in G$ and any two random plans of action (F, π) and $(\tilde{F}, \tilde{\pi})$ such that $(\tilde{F}, \tilde{\pi})$ is a transformation of (F, π) under g. A cost functional C is *invariant* with respect to G if:

$$C(F, \pi) = C(\tilde{F}, \tilde{\pi}) \quad \text{whenever } \tilde{\pi} \text{ is sufficient for } \pi \text{ with respect to } F.$$

Invariance captures the idea that attention costs should not depend on how we describe the world. The sufficiency requirement in the definition ensures we only consider transformations that preserve all information. This means invariance applies when we change descriptions in ways that don't affect the underlying decision problem– like converting prices from dollars to euros, or relabeling productivity states (e.g., from "high" and "low" to "H" and "L").

This property is satisfied by mutual information costs, which depend only on the statistical relationship between actions and states. In fact, a useful (but somewhat informal) heuristic is that a cost function is invariant if and only if it can be expressed as a function of the mutual information between actions and states. The two are closely related (Caplin et al., 2022). On the other hand, invariance rules out attention costs that depend on the variability of consumer prices. We refer the reader to Angeletos and Sastry (2025) and Hébert and La'O (2023) for a deeper discussion of how this definition relates to existing notions of invariance in the literature.

3.2 The irrelevance result in the simple model

Having defined the two key properties of attention costs, we can now state and prove our irrelevance result for the simple two-good economy. Theorem 1 establishes that, under our two assumptions about attention costs, optimal tax formulas take exactly the same form as in classical public finance.

To set the stage, recall that the government's problem is to set taxes in order to raise a given amount of revenue while minimizing distortions. In the standard Ramsey problem with fully attentive agents, the optimal tax in each state θ depends inversely on the price elasticity of demand in that state (Ramsey, 1927). The reason is that the optimal tax system aims to equalize the proportional change in demand across commodities. This implies that the tax burden should concentrate on goods with less elastic demand.[1]

Our result shows that a similar logic applies when agents are rationally inattentive, provided that attention costs satisfy invariance and state separability. The only difference is that we must work with *effective* elasticities rather than the elasticities that would arise under full attention.

To make this precise, we define the effective elasticity of aggregate demand in state θ with respect to the after-tax price state θ':

$$\mathcal{E}(\theta, \theta') \equiv -\frac{d \log \bar{x}(\theta)}{d \log q(\theta')}. \tag{9}$$

This object captures how responsive aggregate consumption is to price changes in a given state, accounting for the effects of this price change on optimal attention. These elasticities summarize all the effects of inattention "under the hood" and hence capture the effective change in demand for good x when the government adjusts taxes. Under invariance and state separability, the own-price elasticity $\mathcal{E}(\theta, \theta)$ is a sufficient statistic for optimal tax design.

Let G^q be the set of functions that transform the after-tax price component of the cognition state: $G^q = \{g : \Theta \times \mathbb{R}_+ \to \Theta \times \mathbb{R}_+ : g(\theta, q) = (\theta, t(q)) \text{ for some } t : \mathbb{R}_+ \to \mathbb{R}_+\}$.

[1] With income effects, one has to use compensated elasticities, but the logic remains the same.

Theorem 1. *If C is invariant with respect to G^q and separable across states, then the optimal tax in state θ satisfies:*

$$\frac{\tau(\theta)}{q(\theta)} = \frac{\Lambda}{\mathcal{E}(\theta,\theta)}, \tag{10}$$

where $\Lambda \equiv 1 - \frac{1}{\lambda}$ is the normalized social value of public funds.

Next, we explain how we prove this result and the roles played by invariance and state separability in each step of the proof. The detailed derivations are in Appendix A.

First, when attention costs are invariant with respect to price transformations, consumer welfare does not depend directly on the stochastic properties of after-tax prices. This means that the indirect utility function can be expressed as a function of the exogenous prior over fundamentals π_Θ, rather than the endogenous prior over cognition states π. More formally, invariance allows us to establish the following result:

Lemma 1. *When C is invariant with respect to G^q, there exists a function $\hat{v}$ such that*

$$v(\boldsymbol{Q}, w, \pi) = \hat{v}(\boldsymbol{Q}, w, \pi_\Theta).$$

This follows directly from Angeletos and Sastry (2025), who show that inattentive economies have an equivalent "twin economy" representation. In this twin economy, reduced-form preferences subsume the attention choice of agents. Under invariance, these preferences depend only on the prior over fundamentals π_Θ. For our purposes, this means that we can ignore the effects of taxes on the prior over cognition states π when computing the effect of taxes on social welfare.

The second step is to notice that state separability breaks any form of interdependence across states. This means that the aggregate demand in state θ can be expressed as a function of the price in that state alone, without reference to prices in other states.

Lemma 2. *If C is separable across states, there exists a function $\hat{F}$ such that for any $\theta \in \Theta$:*

$$F(x \mid z_\theta; \boldsymbol{Q}) = \hat{F}(x \mid z_\theta) \quad \forall x \in \mathcal{X}.$$

In words, state separability implies that the distribution of consumption in any given state θ is independent of the after-tax prices in other states $\theta' \neq \theta$. This follows from inspecting the first-order conditions for the stochastic choice rule F, as detailed in Appendix A.2. This property helps simplify the problem because it means that aggregate demand in state θ can be written as a function of the price in that state alone. This eliminates cross-state interactions that would otherwise complicate the derivation of optimal tax formulas.

Together, these two properties reduce the government's problem to a collection of standard, state-by-state Ramsey problems. Lemma 1 simplifies the first term in the government's objective function (5). Under invariance, the government can ignore the effect of taxes on the endogenous distribution of cognition states π. This helps bypass one of the difficulties of the planning problem with inattentive agents discussed earlier. Lemma 2, on the other hand, simplifies the revenue term in (5). With state separable attention costs, tax revenue in state θ, $\tau(\theta)\bar{x}(\theta)$, depends only on the local after-tax price $q(\theta)$ and not on prices in other states.

These two simplifications imply that the planning problem collapses to a standard Ramsey problem, where taxes affect welfare and tax revenue through the usual channels. Our irrelevance result then follows from the optimality condition for the choice of taxes in each state θ.[2] The standard Ramsey logic yields the familiar inverse-elasticity rule, where inattention affects the outcome only through its impact on the effective elasticity $\mathcal{E}(\theta, \theta)$.

We think this irrelevance result is a useful benchmark that clarifies when and why rational inattention introduces truly novel considerations for optimal taxation. Under our key assumptions, all relevant effects of inattention are captured through effective demand elasticities. Policy makers can therefore proceed as if agents were fully attentive, plugging in measured demand elasticities into familiar optimal tax formulas.

This also shows that the argument for corrective taxation found in some of the existing literature (e.g., Gerritsen, 2016; Allcott et al., 2018; Farhi and Gabaix, 2020) hinges on a paternalistic perspective. When the planner internalizes agents' attention costs, "behavioral wedges" do not enter the optimal tax formulas. Importantly, this result does not imply that inattention is without consequence for tax policy. Equation (10) confirms that the planner should exploit inattention so as to extract a given amount of revenue with less distortion. However, as we show in the application below, inattention does not necessarily make agents less responsive to price changes and hence does not always call for higher taxes.

[2]See Appendix A.1 for the derivation.

3.3 Application to the costly control economy

To make the irrelevance result more concrete, we illustrate its implications using our costly control example from Section 2. This example highlights how inattention can affect the magnitude of demand elasticities while preserving the structure of optimal tax formulas. In addition, it shows that inattention can lead to more elastic demand and may thus call for *lower* taxes.

Consider the costly control example with CRRA utility $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$. Agents face difficulty implementing their desired consumption choices precisely, leading to random deviations from their intended consumption plans. Recall that realized consumption is $\tilde{X} = Xe^{\epsilon}$ where X is the target and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a tremble. We let $k \equiv \sigma_\epsilon^2$ denote the variance of the tremble, which agents choose optimally.

Given k, the optimal target solves the first-order condition and is given by

$$\log X = -\frac{1}{\gamma}\log q + \frac{\gamma-2}{2}k.$$

Aggregate demand in levels is $\bar{X} = \mathbb{E}[Xe^{\epsilon}] = Xe^{k/2}$, so log aggregate demand is

$$\log \bar{X}(\theta) = -\frac{1}{\gamma}\log q(\theta) + \frac{\gamma-1}{2}k(q(\theta)), \tag{11}$$

where $k(q)$ denotes the optimal variance as a function of the after-tax price. The first term is the standard price effect. The second term captures how changes in k affect aggregate consumption. Since k depends on the after-tax price, the effective demand elasticity will typically differ from the fully attentive elasticity γ^{-1}.

Proposition 1 (Costly Control with CRRA)**.** *In the costly control economy with CRRA utility, the optimal tax in state θ solves* (10)*, with the effective elasticity given by:*

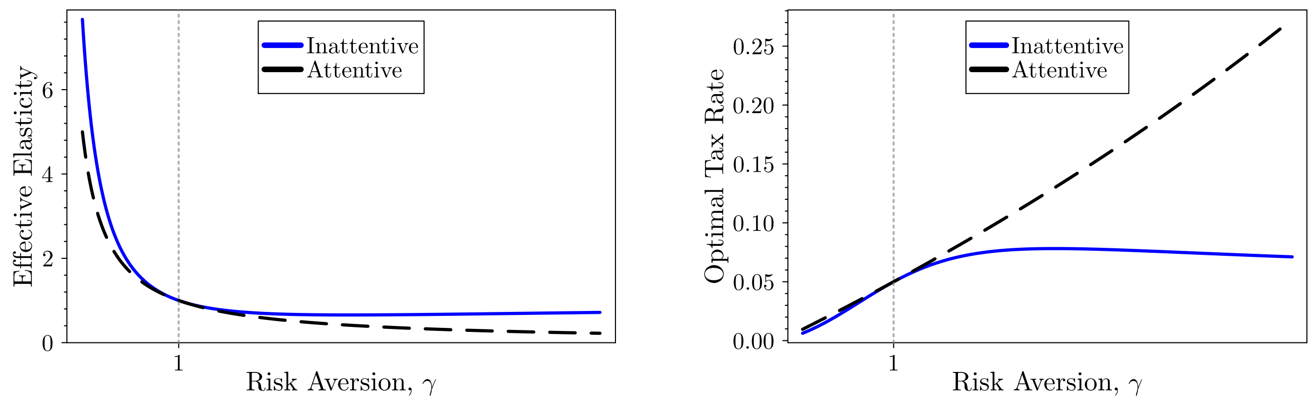
$$\mathcal{E}(\theta,\theta) = \frac{1}{\gamma} - \frac{\gamma-1}{2}\frac{dk(q(\theta))}{d\log q(\theta)} \geq \frac{1}{\gamma}. \tag{12}$$

Proof. See Appendix A.3. □

The first part of the proposition follows from Theorem 1. The second part, proved in Appendix A.3, shows that inattention makes aggregate demand more elastic: the effective elasticity $\mathcal{E}(\theta,\theta)$ is always above the structural elasticity γ^{-1}.

This follows from the comparative static of attention (k) with respect to the after-tax price $q(\theta)$. It turns out that the sign of $\frac{dk(q)}{d\log q}$ depends on whether the risk aversion coefficient γ is above or below 1. If $\gamma > 1$, then attention increases (lower k) as prices increase. If $\gamma < 1$, the opposite holds, but the effective elasticity remains above γ^{-1} because the sign of $\gamma - 1$ also flips. It follows that, regardless of the value of γ, inattention makes agents more responsive and optimal taxes are lower than in the fully attentive case. Figure 2 illustrates these results using an example with an exponential cost function (with parameter κ) that admits a closed-form solution for optimal taxes.[3]

Figure 2: Optimal Tax and Effective Elasticity in Costly Control Economy



(a) Effective Elasticity, $\mathcal{E}(\theta, \theta)$

(b) Optimal Tax Rate (%)

Note: Panel (a) shows the effective elasticity $\mathcal{E}(\theta, \theta)$ as risk aversion γ varies. Panel (b) displays the corresponding optimal tax rates. Both panels compare the case of inattentive agents (solid blue line, with $\kappa = 1.0$) to fully attentive agents (dashed black line).

One way to interpret our comparative static result is to think of attention choice as a kind of "investment". When the after-tax price of the taxable good increases, the rate of return on this "investment" decreases. When $\gamma < 1$, the substitution effect dominates the income effect, so you invest less when the rate of return falls. This means attention decreases (k increases) when the after-tax price increases. When $\gamma > 1$, the income effect dominates the substitution effect, so you invest more as the rate of return falls. In this case, attention increases (k decreases) as prices increase.

In both cases, however, the effective elasticity increases relative to the fully attentive benchmark. The reason is that the relationship between the variance k and aggregate demand also

[3]See Appendix A.3 for details on this example and the derivations.

depends on $\gamma - 1$. To understand why, extend the investment analogy: higher variance lowers the risk-adjusted return on consuming good x. Once again, the response depends on whether the substitution or income effect dominates. When $\gamma < 1$, the substitution effect dominates and aggregate demand falls with variance. When $\gamma > 1$, the income effect dominates and aggregate demand rises with variance.[4] Since the comparative static of attention with respect to prices also depends on the sign of $\gamma - 1$, the two effects reinforce each other: when prices rise, both the change in attention and its effect on demand work in the same direction, guaranteeing that the effective elasticity always exceeds the structural elasticity.

This simple example challenges the view that inattention necessarily makes agents less responsive and thus calls for higher taxes (Mullainathan et al., 2012; Goldin, 2015; Farhi and Gabaix, 2020). Instead, it demonstrates that the effects of inattention on optimal taxes cannot be determined a priori: they depend on the nature of attention costs and how these costs interact with underlying features of the environment. In our costly control example, inattention systematically increases effective elasticities, suggesting that optimal taxes should be *lower* than in the fully attentive case. This highlights the importance of careful empirical work in determining the appropriate elasticities for optimal tax calculations, a point we return to in Section 4.

Having derived the irrelevance result in this specific application, one might wonder whether this result extends beyond the simple quasilinear environment we have considered so far. To address this, we now show that our irrelevance result also holds in a more general model that features multiple goods, heterogeneous agents, and arbitrary utility functions. This extension confirms the robustness of our main result.

3.4 Extension to the general model

This section extends our irrelevance result to a general multi-good economy with ex-ante heterogeneous agents. We show that irrelevance continues to hold, with the only change being that the optimal tax formulas now involve compensated responses and social marginal welfare weights for each agent type. This verifies that our findings are not driven by the restrictions imposed by the simple model of Section 2. Instead, they reflect deeper properties of attention costs that hold more generally. It also connects our results to the literature on optimal commodity

[4] An Epstein-Zin formulation that separates risk aversion from the EIS confirms that it is the EIS, not risk aversion, that governs whether aggregate demand rises or falls with variance. See Appendix A.3 for details.

taxation in economies with heterogeneous agents (Diamond, 1975; Mirrless, 1975).

The model has a similar structure to the one in Section 2 except it allows for an arbitrary number of consumption goods and differences across agents. There is a stochastic state of nature $\theta \in \Theta$, which determines the producer price of N non-contingent goods, indexed by $n \in \{1, \ldots, N\}$. After-tax price of goods $\boldsymbol{q} \in \mathbb{R}^N_+$ is random from the perspective of the agents. The "true" after-tax price of goods in state θ is given by the vector $\boldsymbol{q}(\theta) \equiv \boldsymbol{p}(\theta) + \boldsymbol{\tau}(\theta)$, where $\boldsymbol{p}(\theta)$ is the vector of producer prices and $\boldsymbol{\tau}(\theta)$ is the vector of taxes chosen by the government. Following the simple model, we assume that agents collect information about $z = (\theta, \boldsymbol{q})$ and we refer to z as the cognition state.

There is a continuum of consumers with measure one, split into a finite number of types $j \in \{1, \ldots, J\}$, each with mass μ^j. Each agent type j has its own utility function u^j, endowment w^j, and attention cost function C^j, solving:

$$\begin{aligned} \max_{F^j \in \mathcal{F}} \sum_{z \in Z_\pi} \int_{\mathcal{X}} u^j(x,\theta) dF^j(x \mid z) \pi(z) - C^j(F^j, \pi) \\ \text{s.t.} \quad \int_{\mathcal{X}} \boldsymbol{q} \cdot x \, dF^j(x \mid z) \leq w^j, \quad \forall z \in Z_\pi. \end{aligned} \tag{13}$$

This formulation of the budget constraint for each cognition state z implicitly assumes a strong notion of complete markets. This means agents can trade contingent claims across all states they can distinguish, allowing them to transfer resources and smooth the marginal value of wealth (Angeletos and Sastry, 2025). While somewhat unrealistic, this assumption allows us to separate the effects of inattention from the effects of incomplete markets against idiosyncratic risk.[5] In the absence of a quasilinear good that acts as residual good, it also helps ensure budget constraints hold with equality when agents are inattentive.

The key observation is that invariance and state separability impose the same restrictions on the government's problem, regardless of the number of goods or agents' preferences. Invariance mutes the dependence of indirect utility functions on the endogenous prior π, while state separability continues to ensure that consumption choices in each state depend only on after-tax prices in that state.

Following Diamond (1975), we define the social marginal utility of income for agents of type j in state θ as:

$$\gamma^j(\theta) \equiv \beta^j + \lambda \boldsymbol{\tau}(\theta) \cdot \bar{\boldsymbol{x}}^j_w(\theta), \tag{14}$$

[5] These are well understood by now, see Geanakoplos and Polemarchakis (1986).

where $\beta^j \equiv \frac{\mathcal{W}_{v^j} v^j_w}{\mu^j}$ is the social marginal welfare weight of group j. The definition of $\gamma^j(\theta)$ uses the vector of income effects for group j in state θ $\bar{\boldsymbol{x}}^j_w(\theta) \equiv \left(\frac{\partial \bar{x}^j_1(\theta)}{\partial w}, \ldots, \frac{\partial \bar{x}^j_N(\theta)}{\partial w}\right)^T$. It takes into account that increasing the income of group j in state θ not only increases consumption, valued at β^j, but also generates a "fiscal externality" as some of this income is spent on taxed goods. The term $\lambda \boldsymbol{\tau}(\theta) \cdot \bar{\boldsymbol{x}}^j_w(\theta)$ captures this fiscal externality, priced with the shadow value of public funds λ. This is what drives the difference between social marginal welfare weights and social marginal utilities of income.

We are now ready to state our irrelevance result for the general model.

Theorem 2. *If all agents have attention costs that are invariant and state separable, then the optimal tax on good i in state θ satisfies:*

$$\sum_{j=1}^{J} \mu^j \left[\left\{ \lambda - \gamma^j(\theta) \right\} \bar{x}^j_i(\theta) + \lambda \boldsymbol{\tau}(\theta) \cdot \boldsymbol{S}^j_i(\theta) \right] = 0, \tag{15}$$

*where $\boldsymbol{S}^j_i(\theta)$ denotes the i^{th} column of the **Slutsky matrix** for group j in state θ:*

$$\boldsymbol{S}^j_i(\theta) \equiv \bar{\boldsymbol{x}}^j_{q_i}(\theta) + \bar{\boldsymbol{x}}^j_w(\theta) \bar{x}^j_i(\theta). \tag{16}$$

Proof. See Appendix A.4. □

This result generalizes the Diamond (1975) formulas to economies with rationally inattentive agents. It shows that increasing the tax of good i in state θ affects social welfare through three standard channels. First, it increases tax revenue from group j's consumption of good i by $\bar{x}^j_i(\theta)$. The total value of these additional resources is $\lambda \sum_{j=1}^{J} \mu^j \bar{x}^j_i(\theta)$. Second, it makes consumers poorer, and this lowers welfare by $\sum_{j=1}^{J} \mu^j \gamma^j(\theta) \bar{x}^j_i(\theta)$. Here, $\{\gamma^j(\theta)\}_{j=1}^{J}$ determines the social cost of reducing the income in state θ and thus captures the planner's redistributive motives. Finally, the Slutsky terms capture the distortionary effects of taxation when agents substitute away from taxed commodities. At the optimum, the planner sets taxes to that the sum of these effects cancels out.

This confirms that our irrelevance result holds in a much more general setting than the simple model of Section 2. The only change is that the optimal tax formulas now involve compensated elasticities and social marginal welfare weights, which are standard features of optimal tax formulas in economies with heterogeneous agents. The formulas we obtain are analogous to the ones found in the literature on optimal commodity taxation with heterogenous agents.

In this sense, inattention is irrelevant for optimal tax design as long as attention costs satisfy invariance and state separability.

Note that this result allows for heterogeneity in attention costs across agent types. Different agents may face different costs of paying attention, and these costs may be correlated with other characteristics such as preferences or income. From the perspective of an econometrician estimating optimal taxes, none of these details need to be separately identified: the reduced-form demand responses are sufficient to compute optimal taxes. They capture any confusion between states, the trembles in consumption choices, and the endogenous allocation of attention. Under our assumptions, there is no need to separately estimate or account for these components: they are all summarized by the compensated responses.

This further clarifies an important conceptual point about optimal taxes with inattentive agents. There is no separate "corrective" role for taxes when inattention is the product of rational choice and the planner recognizes this. This is an important difference between the rational inattention approach and approaches that treat inattention as an inherent bias or mistake. Here, as long as invariance and state separability hold, behavioral wedges do not enter the optimality conditions for taxes.

In fact, as we show in the next section, even if we relax state separability, the spirit of optimal tax formulas remains unchanged. Violations of state separability introduce cross-state elasticities: changing taxes in state θ affect demand responses in other states θ', but the core logic remains. The structure of the optimality conditions continues to mirror the classical ones, merely incorporating richer patterns of substitution across states. Relaxing invariance, on the other hand, introduces new considerations that have no analogue in classical public finance. This is where rational inattention really matters for optimal tax design.

4 BEYOND INVARIANCE AND STATE SEPARABILITY

This section explores what happens when we relax the assumptions of invariance and state separability that underlie our irrelevance result. This clarifies the scope of our main result and highlights the new considerations that arise away from this benchmark.

Our analysis proceeds in two steps. We first relax state separability while maintaining invariance. We show that this introduces interdependences across states, meaning that optimal taxes now depend not only on familiar *within-state* elasticities but also on *cross-state* elasticities. While this presents new empirical challenges, it does not change the spirit of optimal tax formulas.

We then consider violations of invariance. This is where rational inattention introduces truly novel considerations for optimal tax design. We find that this opens the door to inefficiencies due to so-called "cognitive externalities". In this environment, taxes take on a new role: they should aim at simplifying the price system. For example, when cognitive costs depend on the variance of after-tax prices, the planner uses taxes as a tool to reduce price volatility. We argue that this perspective helps accommodate the role of salience with rationally inattentive agents. It also helps explain why efficiency and welfare may depend on whether taxes are included in posted prices (as is common in Europe) or added at checkout (as in the United States).

4.1 Relaxing state separability

We begin by deriving the optimal tax formulas when attention costs are invariant but not necessarily state separable. Throughout this section, we work with the general model described in Section 3.4 that features multiple goods and ex-ante heterogeneous agents.

Departures from state separability capture the idea that the cognitive effort required to process information about one state influences, or is influenced by, attention devoted to other states. This is a common feature of rational inattention models, including those with mutual information costs, where an agent's desire to anchor towards commonly played actions links attention choices across states. This, in turn, leads to a situation where consumption choices in any given state depend on the vector of after-tax prices in that particular state and prices in all other states.

Despite this *interdependence*, the key insight is that the structure of the optimal tax formulas remains similar to the ones derived under state separability. The compensated responses, as captured by the Slutsky matrix, and appropriately-defined welfare weights remain sufficient for optimal taxes. The only difference is that the relevant objects now include cross-state interactions.

To see this, define the generalized social marginal utility of income for group j in state θ as:

$$\gamma^j(\theta) \equiv \beta^j + \frac{\lambda}{\pi_\Theta(\theta)} \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \bar{\boldsymbol{x}}^j_w(\theta').$$

Like in Section 3.4, this object captures the social value of increasing the income of group j in state θ after accounting for fiscal externalities. These fiscal externalities now include revenue changes across *all* states $\theta' \in \Theta$ because attention costs are not state separable.

With this, we obtain the following result, generalizing Theorem 2 to the case of invariant but non-separable attention costs:

Theorem 3. *When all agents have attention costs that are invariant with respect to G^q, the optimal tax on good i in state θ satisfies:*

$$\sum_{j=1}^{J} \mu^j \left[\left\{ \lambda - \gamma^j(\theta) \right\} \bar{x}_i^j(\theta) + \frac{\lambda}{\pi_\Theta(\theta)} \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \boldsymbol{S}_i^j(\theta', \theta) \right] = 0, \tag{17}$$

where $\boldsymbol{S}_i^j(\theta', \theta)$ denotes the i^{th} column of the generalized Slutsky matrix for group j:

$$\boldsymbol{S}_i^j(\theta', \theta) = \frac{\partial \bar{\boldsymbol{x}}^j(\theta')}{\partial q_i(\theta)} + \bar{\boldsymbol{x}}_w^j(\theta') \bar{x}_i^j(\theta).$$

Proof. See Appendix B.1. □

This result shows that the optimal taxes can still be written in terms of the classical sufficient statistics even when attention costs are not state separable. The only difference relative to (15) is the presence of the cross-state Slutsky terms $\boldsymbol{S}_i^j(\theta', \theta)$, which capture how consumption in state θ' responds to (compensated) price changes in state θ. These terms reflect the interdependence across states and summarize how changing policy in one state affects behavior in others.

It is worth noting that a similar formula applies in fully attentive economies with preferences that are not separable across states. In that case, the cross-state terms capture departures from expected-utility maximization instead of the "attention spillovers" that arise from violations of state separability. This connection suggests that violations of state separability do not fundamentally alter the structure of optimal tax formulas, but rather introduce non-separabilities that must be accounted for.

To expand on this, we next consider an example with power utility and Gaussian uncertainty.

Log-normal Example

We consider again an economy with quasilinear preferences and two goods: a taxable good X and an untaxable numeraire Y. The exogenous fundamental θ now follows a log-normal distribution, $\log\theta \sim N(0,1)$, and the producer price of the taxable good is given by $p(\theta) = \theta$. The government chooses taxes of the form $\tau(p) = \chi p^{\varphi} - p$, or equivalently an after-tax price

$$q(p) = \chi p^{\varphi},$$

where $p = \theta$ is the realized producer price (in this example also the exogenous fundamental), $\chi > 0$ governs the level of taxation, and φ governs the non-linearity, or the state dependence, in the tax system. We focus on tax systems of this form because they ensure that after-tax prices follow a log-normal distribution: $\log q \sim N(\log\chi, \varphi^2)$

As in our main analysis, agents face a costly information acquisition problem. They do not observe the after-tax price of the taxable good. Instead, they acquire signals of the form $\omega_i = \log(q) + a\epsilon_i$, where $\epsilon_i \sim N(0,1)$ is the noise in the signal and $a \geq 0$ is controlled by the agents. We let $\rho \equiv \frac{\mathbb{V}[\log q]}{a^2+\mathbb{V}[\log q]}$ denote the square of the correlation of the signal with the fundamental and henceforth represent the choice of attention as the choice of ρ. Note this ρ is an a decreasing transformation of a, an increasing transformation of the signal-to-noise ratio, and and an increasing transformation of the entropy between ω and q,[6] so choosing ρ is equivalent to choosing any of these objects.

We finally maximize tractability by letting the cost of attention be proportional as opposed to additive. In particular, we specify the joint preferences over goods and attention as follows:

$$\mathcal{U} = \mathbb{E}\left[U(X,Y)\right] V(\rho),$$

where $U(X,Y) = \frac{1}{1-\gamma}X^{1-\gamma} + Y$ and $V(\rho)$, the cost of attention, will be specified shortly. For now, we note the following: by expressing the cost of attention as a function of ρ alone (or equivalently of the entropy between the signal and the fundamental), we are effectively imposing invariance (as in the case with mutual information costs).

The household budget gives $Y = I - qX$, where I is income. We normalize $I = 0$, which amounts to abstracting from wealth effects in the choice of attention. For any given realization of the signal, the optimal choice of X maximizes $\mathbb{E}\left[U(X,-qX)|\omega\right]$, so the optimal demand is

$$X(\omega) = \left(\mathbb{E}[q|\omega]\right)^{-\eta} = \exp\left\{-\eta\left(\,\rho\omega_i + (1-\rho)\log\mathbb{E}[q]\right)\right\},$$

where $\eta \equiv 1/\gamma$ parameterizes the elasticity of demand. More attention therefore increases, not only the correlation of the signal ω with the price, but also the sensitivity of the optimal

[6]In particular, the signal-to-noise ratio is $\frac{\rho}{1-\rho}$ and the entropy is $-\log(1-\rho)$.

demand on the signal. Using this result, we can show (after some tedious algebra) that

$$\mathbb{E}\left[U(X,Y)\right] = \frac{1}{\eta-1}\exp\left\{\left(1-\tfrac{1}{\eta}\right)\left[-\eta\log\mathbb{E}[q] + \frac{\eta^2\rho}{2}\mathbb{V}(\log q)\right]\right\}.$$

Note that $\mathbb{E}\left[U(X,Y)\right]$ increases with ρ, capturing the benefits of attention. Turning to the cost of attention, we let $V(\rho) = \exp\left\{-(1-\frac{1}{\eta})K(\rho)\right\}$, from some increasing and convex function K.It follows that the optimal ρ solves

$$\max_{\rho\in[0,1]}\left\{\frac{1}{2}\rho\eta^2\mathbb{V}(\log q) - K(\rho)\right\} \tag{18}$$

Optimal attention thus increases with the variance of the (after-tax) price faced by the consumer. Intuitively, the more volatile prices are, the higher the benefit of paying attention to the changes in prices.

When $\varphi = 1$, taxes are state-independent and $\mathbb{V}(\log q) = \mathbb{V}(\log p) = \mathbb{V}(\log\theta) = 1$. More generally, $\mathbb{V}(\log q) = \varphi^2$. By making taxes state-dependent (equivalently, non-linear in p), the planner can thus influence the agents' choice of attention. But is it optimal to do so? And if yes, in what direction? Should the planner increase or reduces the agents' attention relative to the benchmark with state-independent taxes? The next result shows that the answer to this question depends on η, the elasticity of demand.

Proposition 2. *Consider the log-normal example described above. The optimal tax system is state-independent ($\varphi^* = 1$) if and only if $\eta = 1$. If instead the elasticity of demand is $\eta < 1$, then $\varphi^* < 1$: the after-tax price q is concave in p, and equivalently the optimal tax rate is increasing in p. (And the opposite is true if $\eta > 1$.)*

The intuition is as follows. With endogenous attention, the planner benefits from state-dependent taxes because they allow control over how much attention agents pay to prices. By an envelope argument, the first-order effect on individual welfare from distorting attention is zero. What matters is the effect on tax revenue. Starting from a state-independent tax system ($\varphi = 1$), an increase in attention ρ raises tax revenue when $\eta > 1$ and lowers it when $\eta < 1$. When $\eta < 1$, the planner therefore wants to reduce attention so as to increase tax revenue. This is achieved by making after-tax prices less volatile ($\varphi^* < 1$). When $\eta > 1$, the planner wants to increase attention to raise revenue, and does so by making after-tax prices more volatile ($\varphi^* > 1$).

We visualize the result in Figure 3. For this figure, we let the cost of attention be a scalar $\kappa > 0$ times the square of the mutual information between the signal and the price,[7] and we show how the optimal φ^* varies with κ. Although the *sign* of the state-dependence depends merely on whether η is higher or lower than 1, its *magnitude* scales with κ. Intuitively, when κ is close to zero, attention is high and the mechanism described above is weak (indeed it vanishes as $\kappa \to 0$). But once attention is low (which is the case when κ is high), there can be large gains from reallocating the tax distortion across states, in the way explained above.

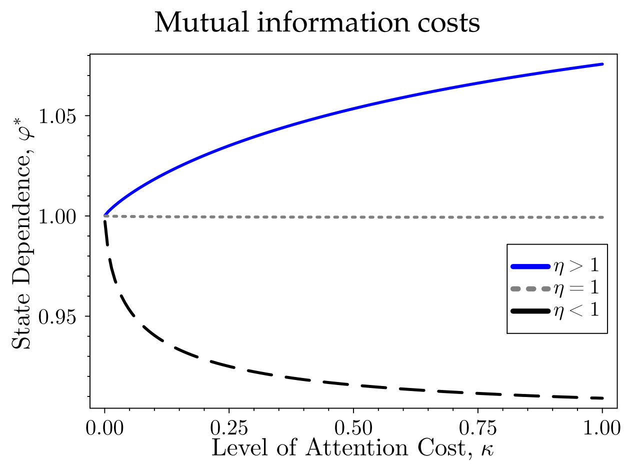


Figure 3: Optimal state dependence due to non-separable costs.

We conclude with two observations. First, the optimality of state-dependent taxes obtained in Proposition 2 depends on the endogeneity of attention: if ρ were exogenously fixed, then the optimal φ would have been 1 regardless of η. Second, by expressing the cost of attention as a function of ρ alone, we have apriori assumed invariance, as in the case with mutual-information costs. The tax properties identified in Proposition 2 thus derive exclusively from the violation of state-separability. We next turn attention to the implications of violating invariance.

4.2 Relaxing invariance

We now turn to violations of invariance. As anticipated above, this is where rational inattention introduces truly novel considerations for optimal taxes. Unlike violations of state separability, which modify the application of familiar principles, departures from invariance introduce what

[7] That is, $K(\rho) = \kappa \times (-\log(1-\rho))^2$.

Angeletos and Sastry (2025) call *cognitive externalities*. These externalities arise when attention costs depend on the volatility or other statistical properties of prices (e.g., their "complexity"), which consumers take as given but the government can influence through its tax policy. This in turn calls for taxes to serve not merely as instruments to raise revenue and redistribute, but also as tools to simplify the price system and thereby reduce the cognitive burden imposed on agents.

To see how these cognitive externalities enter optimal tax formulas, consider the general model of Section 3.4. The next result generalizes Theorem 3 to allow for violations of invariance: the formula retains a similar structure, but a new term appears capturing the welfare effect of a tax change through the induced change in π, the prior over cognition states.

Theorem 4. *When attention costs are neither invariant nor state separable, the optimal tax on good i in state θ satisfies:*

$$\sum_{j=1}^{J} \mu^j \left[\left\{ \lambda - \gamma^j(\theta) \right\} \bar{x}_i^j(\theta) + \frac{\lambda}{\pi_\Theta(\theta)} \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \boldsymbol{S}_i^j(\theta', \theta) + \beta^j \Xi_i^j(\theta) \right] = 0, \tag{19}$$

where $\gamma^j(\theta)$ and $\boldsymbol{S}_i^j(\theta', \theta)$ are as in Theorem 3 and the additional term is

$$\Xi_i^j(\theta) \equiv \frac{1}{v_w^j} \frac{\partial \psi^j}{\partial q_i}(z_\theta), \qquad z_\theta \equiv (\theta, \boldsymbol{q}(\theta)).$$

Here, $v_w^j \equiv \partial v^j / \partial w^j$ and $\psi^j(z)$ is defined (up to an additive constant) by the property that for any perturbation $\Delta\pi$ with $\sum_z \Delta\pi(z) = 0$,

$$\left. \frac{d}{d\varepsilon} v^j(\boldsymbol{Q}, w^j, \pi + \varepsilon \Delta\pi) \right|_{\varepsilon=0} = \sum_z \psi^j(z) \Delta\pi(z).$$

Proof. See Appendix B.2. □

The term $\Xi_i^j(\theta)$ is a *cognitive-externality wedge*. It captures how a marginal change in the after-tax price $q_i(\theta)$ affects welfare by shifting the prior π and hence attention costs.

To build intuition, recall that the prior places probability $\pi_\Theta(\theta)$ on cognition state $z_\theta = (\theta, \boldsymbol{q}(\theta))$. These probability weights are fixed, but when taxes change, the *locations* z_θ shift. Think of $\psi^j(z)$ as how type j's indirect utility changes when the prior shifts towardz (offset by an equal decrease elsewhere so that probabilities still sum to one). Because the support Z_π is finite, we can equivalently interpret $\psi^j(z)$ as a gradient of $v^j(\boldsymbol{Q}, w^j, \pi)$ with respect to the probability vector $(\pi(z))_{z\in Z_\pi}$. The wedge in Theorem 4 is then proportional to how sensitive this gradient is to the after-tax price component of the cognition state.[8]

To illustrate this, we revisit the log-normal example of the previous section, making two changes. First, we impose $\eta = 1$ to shut down the effects of non-separability (Proposition 2), isolating the effects of non-invariance. Second, we introduce non-invariance by letting the costs of attention depend, not only on ρ, but also on the volatility of the underlying prices: instead of $K(\rho)$, the cost of attention is now given by

$$K(\rho, v), \qquad \text{with} \qquad v = \mathbb{V}(\log q).$$

To see a concrete example, suppose that the cost of attention is an increasing function of the reduction in the posterior variance, or equivalently in the entropy of the posterior:

$$\text{cost} = \tilde{K}\left(\mathbb{V}[\log q] - \mathbb{E}\left[\mathbb{V}[\log q \mid \omega]\right]\right),$$

for some increasing and convex function $\tilde{K}$. Since $\mathbb{V}[\log q \mid \omega] = (1-\rho)\mathbb{V}[\log q]$, this example is nested with $K(\rho, v) \equiv \tilde{K}(\rho v)$. For now, we sidestep this specific example and work with a flexible K.

The optimal ρ now solves

$$\max_{\rho\in[0,1]} \left\{\frac{1}{2}\rho\mathbb{V}(\log q) - K(\rho, \mathbb{V}(\log q))\right\} \tag{20}$$

This is the same as (18), except that we have imposed $\eta = 1$ and have changed the cost of attention. We assume that K is such that the solution is always interior and the optimal ρ is increasing in $\mathbb{V}(\log q)$. Since the benefit of attention increases linearly in $\mathbb{V}(\log q)$, the last property will hold as long as the marginal cost of attention decreases or increases less than linearly in $\mathbb{V}(\log q)$. (Recall that the marginal cost was invariant to $\mathbb{V}(\log q)$ in the mutual-information benchmark studied earlier.)

Write the optimal ρ as $\varrho(v) \equiv \arg\max_{\rho\in[0,1]} \left\{\frac{1}{2}\rho v - K(\rho, v)\right\}$ and let $\epsilon(v) \equiv \frac{\partial\varrho(v)}{\partial v}\frac{v}{\varrho}$ denote the elasticity of the optimal attention with respect to the volatility in prices. In equilibrium, $v = \varphi^2\mathbb{V}(\log\theta)$. This highlights how the planner can regulate the attention by choosing appropriate the state-contingency of the taxes, or equivalently the variability of the after-tax prices faced by

[8] The derivative $\frac{\partial\psi^j}{\partial q_i}$ is taken with respect to the price component of the cognition state, holding the actual after-tax price fixed.

the consumer. The next result then show how exactly the planner leverages this instrument along the optimum.

Proposition 3. *Consider the log-linear economy, with $\eta = 1$ and K as described above. The optimal tax system satisfies*

$$1 - \varphi^* \propto \left. \frac{K_v(\rho(v), v)}{1 + \epsilon(v)} \right|_{v = (\varphi^*)^2 \mathbb{V}(\log \theta)}.$$

That is, the optimal tax rate increases (respectively, decreases) in p if and only if the cost of attention increases (respectively, decreases) in the volatility of prices.

To understand this result, recall that $\eta = 1$ shuts down the effect of non-separability seen in the previous section: there is no scope for manipulating attention in the hope of raising more tax revenue. By the same token, $\eta = 1$ guarantees that the optimal tax system is state-independent when attention costs are invariant—which herein maps to $K_v = 0$, that is, no cognitive externality from the variability of prices. But once such an externality is present ($K_v \neq 0$), a new rationale for manipulating prices and attention choices emerges: such manipulation can improve welfare by "easing" the agents' attention problem. Proposition 3 shows how exactly this rationale translates in terms of the optimal state-contingency of the price system: the optimal tax rate must increase with p if the externality is negative, and decrease with p if the externality is positive.

Proposition 3 further show that the *intensity* of the optimal state-contingency is inversely related to the elasticity of attention: the more elastic attention is, the lower the absolute sensitivity of the optimal tax rate with respect to p (i.e., the closer φ^* is to 1, on either side). This reflects the trade off between two goals: easing the externality ("Pigou") and minimizing the tax distortion of raising tax revenue ("Ramsey"). By assuming $\eta = 1$, we have made sure that the latter goal is achieved with $\varphi = 1$ and $\rho = \varrho(1)$. To ease the externality, the planner must move away from this point in the direction of either $\varphi < 1$ and $\rho < \varrho(1)$ or $\varphi > 1$ and $\rho > \varrho(1)$. The more elastic attention is, the higher the associated deadweight loss of such a movement, and hence also the optimal size of this movement.

We further illustrate the effect of non-invariance in Figure 4, using the example with "variance-reduction costs" mentioned earlier: we let attention costs be a scalar $\kappa > 0$ times the square of the reduction in posterior variance.[9] This case is not directly nested in Proposition 3, because

[9]That is, the cost is $K(\rho, v) = \kappa \times (\mathbb{V}[\log q] - \mathbb{E}[\mathbb{V}[\log q \mid \omega]])^2 = \kappa(\rho v)^2$.

the optimal ρ hits the corner 1 when v is small enough and becomes decreasing in v for v large enough. Still, as the figure shows, the essence remains unchanged.[10] The dashed line corresponds to $\eta = 1$ and isolates the effect of non-invariance; in line with Proposition 3, the optimal tax system has $\varphi^* < 1$. The two other lines let $\eta \neq 1$, thus interacting the effect of non-invariance with that of non-separability (from Proposition 2). When $\eta < 1$, the two forces work in the same direction; when $\eta > 1$, they work in opposite direction. In all cases, the magnitude of the state-dependence increases with κ (and of course vanishes when $\kappa \to 0$) .

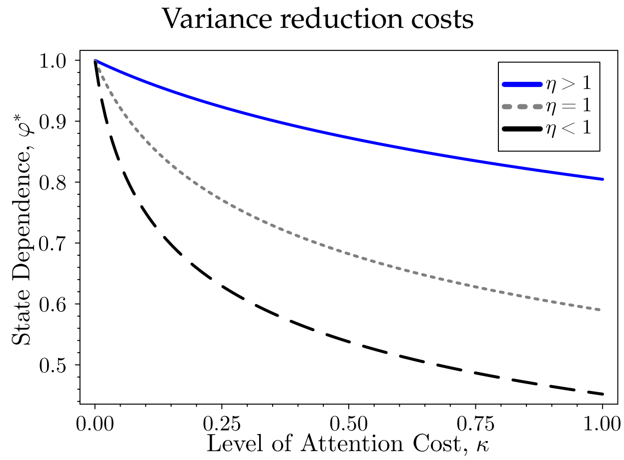


Figure 4: Optimal state dependence φ^* with non-invariant costs.

We conclude with the following remark. Suppose that the planner could use not only the tax system allowed in our paper but also directly control the consumers' attention choices and their demand schedules. In this case, the planner could achieve significantly better outcomes, potentially even the first best. Consider in particular the example with variance-reduction costs. By letting $\varphi \approx 0$ (equivalently, letting q be an exceedingly flat function of the underlying fundamental θ), the planner could virtually eliminate attention costs, even when consumers choose $\rho = 1$. At the same time, by dictating consumers to follow a demand schedule that is extremely steep in their signals (and thereby in q), the planner could implement nearly the same allocation as in the first best. This, by contrast, is infeasible here precisely because the planner cannot freely regulate demand: under laissez faire, consumers have no incentive to vary their demand unless q itself varies enough. It follows, in our context, that the optimal φ balances the value of

[10] In fact, Proposition 3 remains intact if we replace the assumption that ρ is increasing in v with the weaker assumption that attention does not decrease too fast with v, in the sense of $\epsilon(v) > -1$.

easing the cognitive externality against the allocative role of the price mechanism.

Connections to behavioral public finance

Our analysis of non-invariant attention costs connects to several themes in behavioral public finance.

First, it provides a microfoundation for the role of *salience* in optimal tax design. A key insight from that literature is that how taxes are presented can affect behavior and welfare (Chetty et al., 2009). In our framework, this corresponds to whether attention costs depend on the properties of the price distribution. When they do (non-invariance), the distinction between tax-inclusive pricing (common in Europe) and tax-exclusive pricing (common in the US) becomes welfare-relevant. Under invariant costs, this distinction has no effect: the cognitive burden is the same regardless of how prices are quoted. Under non-invariant costs, however, Proposition 3 shows that the planner should actively use taxes to manage price volatility, with the optimal design reflecting a trade-off between easing cognitive costs (Pigou) and minimizing revenue distortions (Ramsey).

Second, our results formalize the idea that tax systems can be *too complex*. In our framework, complexity maps to high variance in after-tax prices. When attention costs violate invariance, the planner faces a trade-off: more volatile prices convey more information but also impose higher cognitive costs. Theorem 4 shows that the optimal tax system must account for this by including a cognitive-externality wedge that has no analog in classical public finance.

5 Conclusion

This paper studies optimal commodity taxation when consumers are rationally inattentive. We embed a flexible form of rational inattention into the classic Ramsey problem and characterize when and how inattention matters for optimal tax policy.

Our main result identifies conditions under which inattention is irrelevant for optimal taxes. Specifically, we show that when attention costs satisfy two key properties— invariance and state separability— optimal taxes satisfy exactly the same sufficient-statistics formula as in classical public finance. Under these conditions, policy makers can totally ignore the presence of inattention and proceed as if agents were fully attentive.

This irrelevance result provides an important benchmark for three reasons. First, it clarifies that arguments for corrective taxation found in some behavioral public finance literature hinge on a paternalistic perspective that treats inattention as an inherent distortion. When the planner recognizes that inattention is the product of rational choice, behavioral wedges do not enter the optimality conditions for taxes.

Second, it qualifies the validity of the argument for exploiting people's inattention for the sake of raising tax revenue with less distortion. Importantly, inattention does not necessarily make agents less responsive to taxes. In some cases, it can *increase* effective demand elasticities, implying that optimal taxes should be *lower* than in the fully attentive case.

Finally, it helps organize the novel considerations that arise when the benchmark assumptions are violated. Relaxing state separability preserves the spirit of optimal tax formulas while introducing measurement challenges. When attention costs are not state separable, the effective elasticities incorporate cross-state interactions but they remain sufficient for optimal tax policy. Conceptually, this is similar to a fully attentive economy with preferences that are not separable across states.

Departures from invariance introduce truly novel considerations. When attention costs depend on the statistical properties of prices, the government can influence agents' cognitive burden. This gives rise to cognitive externalities: individual agents take the price distribution as given, but the government can shape it through tax policy. Optimal taxes then balance the familiar goal of minimizing revenue distortions against a new Pigouvian motive to ease the cognitive burden on consumers.

A Proofs for Section 3

This appendix contains proofs for Section 3. Appendix A.1 completes the proof of Theorem 1, showing that optimal taxes follow a standard inverse elasticity rule whenever attention costs are invariant and separable across states. Appendix A.2 proves Lemma 2, which shows that state separability implies that the distribution of consumption in each state depends only on the after-tax price in that state, not on prices in other states.

A.1 Proof of Theorem 1

The government's problem is to choose taxes $\tau(\theta)$ for each state $\theta \in \Theta$ to maximize (5). Using Lemma 1, the agent's indirect utility can be written as $\hat{v}(\boldsymbol{Q}, w, \pi_\Theta)$, where $\boldsymbol{Q}$ is the stacked vector of after-tax prices $\{q(\theta)\}_{\theta\in\Theta}$, w is wealth, and π_Θ is the prior over fundamental states.

Note that since $q(\theta) = p(\theta) + \tau(\theta)$ and producer prices $p(\theta)$ are treated as exogenous, choosing the after-tax price $q(\theta)$ is equivalent to choosing the tax $\tau(\theta)$. We can therefore think of the government as directly choosing after-tax prices to solve the following problem:

$$\max_{\{q(\theta)\}_{\theta\in\Theta}} \hat{v}(\boldsymbol{Q}, w, \pi_\Theta) + \lambda \left\{ \sum_{\theta\in\Theta} \pi_\Theta(\theta)(q(\theta) - p(\theta))\bar{x}(\theta) \right\}$$

By Lemma 2 (see Appendix A.2 below), $\bar{x}(\theta)$ depends only on the after-tax price in that state. Therefore, the first-order condition with respect to $q(\theta)$ is:

$$\frac{\partial \hat{v}}{\partial q(\theta)} + \lambda \pi_\Theta(\theta) \left\{ \bar{x}(\theta) + (q(\theta) - p(\theta)) \frac{\partial \bar{x}(\theta)}{\partial q(\theta)} \right\} = 0; \tag{21}$$

Using Roy's Identity, $\frac{\partial \hat{v}}{\partial q(\theta)} = -\alpha^m \pi_\Theta(\theta)\bar{x}(q(\theta))$, where $\alpha^m \equiv \frac{\partial \hat{v}}{\partial w}$ is the private marginal utility of income. With quasilinear preferences, $\alpha^m = 1$. Substituting into (21) and simplifying:

$$-\bar{x}(\theta) + \lambda \left\{ \bar{x}(\theta) + \tau(\theta) \frac{\partial \bar{x}(\theta)}{\partial q(\theta)} \right\} = 0$$

Dividing through by $\lambda\bar{x}(\theta)$, rearranging, and using the definition of the effective own-price elasticity $\mathcal{E}(\theta, \theta) \equiv -\frac{d \log \bar{x}(\theta)}{d \log q(\theta)}$ from the main text:

$$1 - \frac{\tau(\theta)}{q(\theta)} \mathcal{E}(\theta, \theta) = \frac{1}{\lambda}$$

Solving for $\frac{\tau(\theta)}{q(\theta)}$ and using $\Lambda \equiv 1 - \frac{1}{\lambda}$, we arrive to equation (10) in the main text:

$$\frac{\tau(\theta)}{q(\theta)} = \frac{\Lambda}{\mathcal{E}(\theta, \theta)}.$$

A.2 Proof of Lemma 2

The individual problem, as defined in Section 2, is to choose a state-dependent stochastic choice rule $F = \{F(x \mid z)\}_{z \in Z_\pi}$ to maximize:

$$\sum_{z \in Z_\pi} \int_{\mathcal{X}} U(x,z) dF(x \mid z) \pi(z) - C(F, \pi),$$

where $U(x,z) \equiv u(x,\theta) + w - qx$ denotes utility in cognition state z after substituting in the budget constraint.[11]

For simplicity, we will focus on the case where F admits densities $f(x \mid z)$, so that $F(x \mid z) = \int_{-\infty}^{x} f(s \mid z) ds$. Recall that the the cost functional $C(F, \pi)$ is state separable if it takes the form:

$$C(F, \pi) = \sum_{z \in Z_\pi} \mu(z) \pi(z) \int_{\mathcal{X}} \phi(f(x \mid z)) dx,$$

where ϕ is a strictly convex function and $\mu(z) > 0$ is a weighting function. Substituting this into the objective function, the problem becomes:

$$\max_{\{f(\cdot|z)\}_{z \in Z_\pi}} \sum_{z \in Z_\pi} \pi(z) \left\{ \int_{\mathcal{X}} U(x,z) f(x \mid z) dx - \mu(z) \int_{\mathcal{X}} \phi(f(x \mid z)) dx \right\},$$

subject to $\int_{\mathcal{X}} f(x \mid z) dx = 1$ and $f(x \mid z) \geq 0$ for all x, z.

This problem can be solved state-by-state. For each $z \in Z_\pi$, the agent chooses the density $f(\cdot \mid z)$ to maximize:

$$\int_{\mathcal{X}} U(x,z) f(x \mid z) dx - \mu(z) \int_{\mathcal{X}} \phi(f(x \mid z)) dx,$$

subject to $\int_{\mathcal{X}} f(x \mid z) dx = 1$. Ignoring the non-negativity constraint, the Lagrangian for this subproblem is:

$$\mathcal{L}_z = \int_{\mathcal{X}} \left[U(x,z) f(x \mid z) - \mu(z) \phi(f(x \mid z)) \right] dx - \nu(z) \left(\int_{\mathcal{X}} f(x \mid z) dx - 1 \right),$$

where $\nu(z)$ is the Lagrange multiplier for state z that ensures $f(x \mid z)$ is a proper density. The first-order condition with respect to $f(x \mid z)$ is:

$$U(x,z) - \mu(z) \phi'(f(x \mid z)) - \nu(z) = 0.$$

Solving for $f(x \mid z)$:

$$f(x \mid z) = (\phi')^{-1} \left(\frac{U(x,z) - \nu(z)}{\mu(z)} \right).$$

[11] We suppress the dependence of U on w as this is held fixed throughout the analysis.

The multiplier $\nu(z)$ is pinned-down by the constraint $\int_{\mathcal{X}} f(x \mid z)dx = 1$. Thus, the optimal $f(x \mid z)$ is entirely determined by cognition state $z = (\theta, q)$ and does not depend on the distribution of actions chosen in any other state $z' \neq z$. The optimal cumulative distribution is $F(x \mid z) = \int_{-\infty}^{x} f(s \mid z)ds$.

In the context of the government's problem, the relevant cognition state associated with a fundamental state θ is $z_\theta = (\theta, q(\theta))$. The agent's chosen consumption distribution in this situation is precisely $F(x \mid z)$ evaluated at $z = z_\theta$. Since $F(x \mid z)$ depends only on z, it follows that we can write $F(x \mid z_\theta; \mathbf{Q}) = \hat{F}(x \mid z_\theta)$, where the function $\hat{F}$ depends on $z_\theta = (\theta, q(\theta))$ and potentially w, but not on $q(\theta')$ for $\theta' \neq \theta$. This proves the lemma.

A.3 Proof of Proposition 1

Proposition 1 states that in the costly control economy with CRRA utility, the optimal tax in state θ solves (10), with the effective elasticity given by:

$$\mathcal{E}(\theta, \theta) = \frac{1}{\gamma} - \frac{\gamma - 1}{2} \frac{d\sigma_\epsilon^2(q(\theta))}{d \log q(\theta)}.$$

The first part of the proposition, that the optimal tax solves (10), follows directly from Theorem 1, as the costly control economy satisfies invariance and state separability.

We need to prove the second part, namely that $\mathcal{E}(\theta, \theta) \geq \frac{1}{\gamma}$. This inequality is equivalent to showing that the term $S \equiv -(\gamma - 1)\frac{d\sigma_\epsilon^2(q(\theta))}{d \log q(\theta)}$ is non-negative. To show this, we turn to the agent's decision problem for choosing the optimal variance of trembles, $k(q) \equiv \sigma_\epsilon^2(q)$.

In the costly control economy with CRRA utility $u(x) = x^{1-\gamma}/(1-\gamma)$, the agent chooses, in each state, a target consumption level X and a variance k to maximize expected utility net of control costs. Given an after-tax price q, the agent's objective is:

$$\max_{X, k \geq 0} \quad \mathbb{E}_\epsilon \left[\frac{(Xe^\epsilon)^{1-\gamma}}{1-\gamma} - qXe^\epsilon \right] - C(k)$$

where $\epsilon \sim \mathcal{N}(0, k)$, and $C(k)$ is the cost of choosing variance k. We assume $C(k)$ is decreasing, reflecting that lower variance (higher precision) is more costly. The expectation can be evaluated as $\frac{X^{1-\gamma}}{1-\gamma} e^{(1-\gamma)^2 k/2} - qXe^{k/2}$. Optimizing with respect to X yields the optimal $X^*(k, q) = q^{-1/\gamma} e^{-k(2-\gamma)/2}$. Substituting X^* back into the objective function, the problem reduces to choosing k to maximize:

$$W(k, q) = \frac{\gamma}{1-\gamma} q^{(\gamma-1)/\gamma} e^{\frac{k}{2}(\gamma - 1)} - C(k)$$

Let $k^*(q)$ denote the optimal choice of variance that maximizes $W(k,q)$. To determine how $k^*(q)$ changes with q, we examine the cross-partial derivative $\frac{\partial^2 W(k,q)}{\partial k \partial q}$. According to standard monotone comparative statics results (e.g., Topkis's Theorem), the sign of this derivative determines the monotonicity of $k^*(q)$:

$$\frac{\partial^2 W(k,q)}{\partial k \partial q} = -\left(\frac{\gamma-1}{2}\right) q^{-1/\gamma} e^{\frac{k}{2}(\gamma-1)}$$

For $\gamma \neq 1$, we have $\frac{\partial^2 W}{\partial k \partial q} < 0$ if and only if $\gamma > 1$. Therefore, $k^*(q)$ is decreasing in q when $\gamma > 1$ and increasing when $\gamma < 1$. It follows that

$$(\gamma - 1)\frac{dk^*(q)}{d \log q} \leq 0 \quad \text{for all } \gamma > 0,$$

with equality if and only if $\gamma = 1$. This immediately implies $S = -(\gamma-1)\frac{dk^*(q)}{d \log q} \geq 0$, and hence $\mathcal{E}(\theta,\theta) = \frac{1}{\gamma} + \frac{S}{2} \geq \frac{1}{\gamma}$.

EXAMPLE WITH EXPONENTIAL COST FUNCTION

Suppose the cost function is of the form $C(k) = e^{-\kappa k}$ for some $\kappa > 0$. To ensure the objective $W(k,q)$ is strictly concave, we assume $\gamma > 1 - 2\kappa$. The first-order condition for optimal k is:

$$-\frac{\gamma}{2} q^{(\gamma-1)/\gamma} e^{\frac{k}{2}(\gamma-1)} + \kappa e^{-\kappa k} = 0$$

Solving for k, we obtain the optimal variance $k^*(q)$:

$$k^*(q) = \frac{2}{2\kappa + \gamma - 1} \max\left\{\ln\left(\frac{2\kappa}{\gamma}\right) - \frac{\gamma-1}{\gamma}\ln(q), 0\right\}.$$

Given this, the effective elasticity is

$$\mathcal{E}(\theta,\theta) = \frac{1}{\gamma} + \frac{(\gamma-1)^2}{\gamma(2\kappa+\gamma-1)}. \tag{22}$$

Once again, we see that $\mathcal{E}(\theta,\theta) \geq \frac{1}{\gamma}$ for all $\gamma > 0$, with equality only when $\gamma = 1$. What's special about this example is that $\mathcal{E}(\theta,\theta)$ is independent of the after-tax price and hence we can read off the optimal tax directly from (10).

A.3.1 EPSTEIN-ZIN PREFERENCES

Suppose risk is aggregated via CRRA aggregator with coefficient $\gamma > 0$, and the intratemporal aggregator is isoelastic with curvature $\rho > 0$ (EIS $= 1/\rho$):

$$H(x) = \frac{x^{1-\rho}}{1-\rho}, \qquad H'(x) = x^{-\rho}. \tag{23}$$

For the multiplicative tremble $\tilde{X} = Xe^{\varepsilon}$ with $\varepsilon \sim \mathcal{N}(0, k)$, the CRRA certainty equivalent of $\tilde{X}$ is

$$\mathrm{CE}_{\gamma}(\tilde{X}) = \Big(\mathbb{E}[\tilde{X}^{1-\gamma}]\Big)^{1/(1-\gamma)} = X\, e^{(1-\gamma)k/2}. \tag{24}$$

Maximizing $H(\mathrm{CE}_{\gamma}(\tilde{X})) - q\,\mathbb{E}[\tilde{X}]$ yields the FOC

$$X^{-\rho}\, e^{(1-\gamma)(1-\rho)\,k/2} = q\, e^{k/2}. \tag{25}$$

Aggregate demand in levels remains $\bar{X} = \mathbb{E}[\tilde{X}] = Xe^{k/2}$, so

$$\log \bar{X}(q,k) = \log X + \frac{k}{2} = -\frac{1}{\rho}\log q + \Big(\frac{\gamma\rho - \gamma - \rho}{2\rho} + \frac{1}{2}\Big)k \tag{26}$$

$$= -\frac{1}{\rho}\log q - \frac{\gamma(1-\rho)}{2\rho}\,k. \tag{27}$$

Thus the effect on aggregate demand depends only on the EIS parameter ρ:

- If $\rho < 1$ (EIS > 1), aggregate demand *falls* with k.
- If $\rho = 1$ aggregate log demand does not depend on k.
- If $\rho > 1$ (EIS < 1), aggregate demand *increases* with k.

A.4 Proof of Theorem 2

We follow the same approach as in the proof of Theorem 1, but now extended to handle multiple goods and heterogeneous agent types. The key insight is that Lemma 1 applies directly to each agent type, while Lemma 2 can be extended to the multi-good setting with budget constraints.

The government's problem is to choose a collection of tax vectors $\{\boldsymbol{\tau}(\theta)\}_{\theta\in\Theta}$ to maximize:

$$\max_{\{\boldsymbol{\tau}(\theta)\}} \mathcal{W}\left(\left\{\hat{v}^j(\boldsymbol{Q}, w^j, \pi_\Theta)\right\}_{j=1}^{J}\right) + \lambda \sum_{j=1}^{J} \mu^j \sum_{\theta\in\Theta} \pi_\Theta(\theta)\boldsymbol{\tau}(\theta)\cdot\bar{\boldsymbol{x}}^j(\theta),$$

where $\hat{v}^j(\boldsymbol{Q}, w^j, \pi_\Theta)$ is agent j's indirect utility, $\boldsymbol{Q}$ is the collection of all after-tax price vectors $\{\boldsymbol{q}(\theta)\}_{\theta\in\Theta}$, μ^j is the population share of type j, and λ is the shadow value of tax revenue. Since producer prices $\boldsymbol{p}(\theta)$ are exogenous, choosing the tax vector $\boldsymbol{\tau}(\theta) = \boldsymbol{q}(\theta) - \boldsymbol{p}(\theta)$ is equivalent to choosing the after-tax price vector $\boldsymbol{q}(\theta)$. The first-order condition with respect to $q_k(\theta)$, the after-tax price of good k in state θ, is:

$$\sum_{j=1}^{J}\frac{\partial\mathcal{W}}{\partial\hat{v}^j}\frac{\partial\hat{v}^j}{\partial q_k(\theta)} + \lambda\sum_{j=1}^{J}\mu^j\pi_\Theta(\theta)\left(\bar{x}_k^j(\theta) + \sum_{n=1}^{N}(q_n(\theta) - p_n(\theta))\frac{\partial\bar{x}_n^j(\theta)}{\partial q_k(\theta)}\right) = 0.$$

Using Roy's Identity, $\frac{\partial \hat{v}^j}{\partial q_k(\theta)} = -\hat{v}^j_w \pi_\Theta(\theta)\bar{x}^j_k(\theta)$, where $\hat{v}^j_w \equiv \frac{\partial \hat{v}^j}{\partial w^j}$ denotes the marginal utility of wealth for agent j. Substituting this into the FOC, and noting that $\tau_n(\theta) = q_n(\theta) - p_n(\theta)$:

$$\sum_{j=1}^{J} \left(\frac{\partial \mathcal{W}}{\partial \hat{v}^j} (-\hat{v}^j_w \pi_\Theta(\theta)\bar{x}^j_k(\theta)) \right) + \lambda \sum_{j=1}^{J} \mu^j \pi_\Theta(\theta) \left(\bar{x}^j_k(\theta) + \sum_{n=1}^{N} \tau_n(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial q_k(\theta)} \right) = 0.$$

Use the definition of social marginal welfare weight for type j, $\beta^j \equiv \frac{1}{\mu^j} \frac{\partial \mathcal{W}}{\partial \hat{v}^j} \hat{v}^j_w$ and divide through by $\pi_\Theta(\theta)$. The FOC becomes:

$$\sum_{j=1}^{J} \mu^j \left(-\beta^j \bar{x}^j_k(\theta) + \lambda \left(\bar{x}^j_k(\theta) + \sum_{n=1}^{N} \tau_n(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial q_k(\theta)} \right) \right) = 0,$$

which can be rewritten as:

$$\sum_{j=1}^{J} \mu^j \left((\lambda - \beta^j)\bar{x}^j_k(\theta) + \lambda \sum_{n=1}^{N} \tau_n(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial q_k(\theta)} \right) = 0. \tag{28}$$

Recall the Slutsky equation: $\frac{\partial \bar{x}^j_n(\theta)}{\partial q_k(\theta)} = S^j_{nk}(\theta) - \bar{x}^j_k(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial w^j}$, where $S^j_{nk}(\theta)$ is the (n,k)-th element of the Slutsky matrix for agent j in state θ. Substitute the Slutsky equation into the sum involving tax terms in (28):

$$\begin{aligned} \sum_{n=1}^{N} \tau_n(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial q_k(\theta)} &= \sum_{n=1}^{N} \tau_n(\theta) \left(S^j_{nk}(\theta) - \bar{x}^j_k(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial w^j} \right) \\ &= \sum_{n=1}^{N} \tau_n(\theta) S^j_{nk}(\theta) - \bar{x}^j_k(\theta) \sum_{n=1}^{N} \tau_n(\theta) \frac{\partial \bar{x}^j_n(\theta)}{\partial w^j}. \end{aligned}$$

Let $\boldsymbol{S}^j(\theta)$ be the Slutsky matrix for agent j in state θ. The first summation $\sum_{n=1}^{N} \tau_n(\theta) S^j_{nk}(\theta)$ is $\boldsymbol{\tau}(\theta) \cdot \boldsymbol{S}^j_k(\theta)$ from Theorem 2 in the main text, where $\boldsymbol{S}^j_k(\theta)$ is the k^{th} column of $\boldsymbol{S}^j(\theta)$ (i.e., its elements are $S^j_{nk}(\theta)$ for $n = 1, \ldots, N$). Now, let $\bar{\boldsymbol{x}}^j_w(\theta)$ denote the vector of marginal responses of demands to wealth for agent j in state θ, i.e., $\bar{\boldsymbol{x}}^j_w(\theta) = \left(\frac{\partial \bar{x}^j_1(\theta)}{\partial w^j}, \ldots, \frac{\partial \bar{x}^j_N(\theta)}{\partial w^j} \right)^T$. Rewriting (28) using vector notation and rearranging terms:

$$\sum_{j=1}^{J} \mu^j \left((\lambda - \beta^j - \lambda \boldsymbol{\tau}(\theta) \cdot \bar{\boldsymbol{x}}^j_w(\theta))\bar{x}^j_k(\theta) + \lambda \boldsymbol{\tau}(\theta) \cdot \boldsymbol{S}^j_k(\theta) \right) = 0.$$

Using the definition $\gamma^j(\theta) \equiv \beta^j + \lambda \boldsymbol{\tau}(\theta) \cdot \bar{\boldsymbol{x}}^j_w(\theta)$ from the main text (see Theorem 2), we arrive at the expression in Theorem 2:

$$\sum_{j=1}^{J} \mu^j \left[\left\{ \lambda - \gamma^j(\theta) \right\} \bar{x}^j_k(\theta) + \lambda \boldsymbol{\tau}(\theta) \cdot \boldsymbol{S}^j_k(\theta) \right] = 0.$$

This completes the proof.

B Proofs for Section 4

This appendix contains the proofs for the results presented in Section 4 of the main text. Appendix B.1 provides a detailed derivation for Theorem 3, which characterizes the optimal taxes when attention costs violate the separability assumption.

B.1 Proof of Theorem 3

With non-separable attention costs, the first order condition with respect to the price of good i in state θ becomes:

$$\sum_{j=1}^{J} \mu^j \left[-\beta^j \pi_\Theta(\theta) \bar{x}_i^j(\theta) + \lambda \pi_\Theta(\theta) \bar{x}_i^j(\theta) + \lambda \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \sum_{n=1}^{N} \tau_n(\theta') \frac{\partial \bar{x}_n^j(\theta')}{\partial q_i(\theta)} \right] = 0. \tag{29}$$

The difference from the separable case analyzed in Appendix A.4 is that changes in the price of good i in state θ affect the demand for goods in *all* states $\theta' \in \Theta$.

The generalized Slutsky equation is:

$$\frac{\partial \bar{x}_n^j(\theta')}{\partial q_i(\theta)} = S_{ni}^j(\theta', \theta) - \frac{\partial \bar{x}_n^j(\theta')}{\partial w^j} \bar{x}_i^j(\theta),$$

where $S_{ni}^j(\theta', \theta)$ is the (n, i)-th element of the generalized Slutsky matrix for agent j, capturing the change in compensated demand of good n in state θ' to a price change of good i in state θ. Substituting this into the last term of (29):

$$\sum_{n=1}^{N} \tau_n(\theta') \frac{\partial \bar{x}_n^j(\theta')}{\partial q_i(\theta)} = \sum_{n=1}^{N} \tau_n(\theta') S_{ni}^j(\theta', \theta) - \left(\sum_{n=1}^{N} \tau_n(\theta') \frac{\partial \bar{x}_n^j(\theta')}{\partial w^j} \right) \bar{x}_i^j(\theta)$$

Plugging this back into (29) and using the vector notation in the main text:

$$\sum_{j=1}^{J} \mu^j \left[(\lambda - \beta^j) \pi_\Theta(\theta) \bar{x}_i^j(\theta) + \lambda \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \left\{ \boldsymbol{\tau}(\theta') \cdot \boldsymbol{S}_i^j(\theta', \theta) - \boldsymbol{\tau}(\theta') \cdot \bar{\boldsymbol{x}}_w^j(\theta') \bar{x}_i^j(\theta) \right\} \right] = 0,$$

where $\boldsymbol{\tau}(\theta')$ is the vector of taxes in state θ' and $\boldsymbol{S}_i^j(\theta', \theta)$ is the i^{th} column of the generalized Slutsky matrix of type j agents. Collecting terms:

$$\sum_{j=1}^{J} \mu^j \left[\left\{ (\lambda - \beta^j) \pi_\Theta(\theta) - \lambda \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \bar{\boldsymbol{x}}_w^j(\theta') \right\} \bar{x}_i^j(\theta) \right.$$

$$\left. + \lambda \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \boldsymbol{S}_i^j(\theta', \theta) \right] = 0$$

Dividing through by $\pi_\Theta(\theta)$ and using the definition of $\gamma^j(\theta)$ in Section 4, we arrive to the expression in Theorem 3:

$$\sum_{j=1}^{J} \mu^j \left[\left\{ \lambda - \gamma^j(\theta) \right\} \bar{x}_i^j(\theta) + \frac{\lambda}{\pi_\Theta(\theta)} \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \boldsymbol{\tau}(\theta') \cdot \boldsymbol{S}_i^j(\theta', \theta) \right] = 0.$$

This completes the derivation.

B.2 Proof of Theorem 4

This proof parallels Theorem 3, except that the indirect utility v^j now depends on the prior over cognition states, which is pinned down by the tax policy and therefore varies with the after-tax price system $\boldsymbol{Q}$.

Let $\pi(\cdot; \boldsymbol{Q})$ denote the prior and define $\tilde{v}^j(\boldsymbol{Q}) \equiv v^j(\boldsymbol{Q}, w^j, \pi(\cdot; \boldsymbol{Q}))$. Because the prior depends on $\boldsymbol{Q}$, the chain rule gives

$$\frac{d\tilde{v}^j}{dq_i(\theta)} = \frac{\partial v^j}{\partial q_i(\theta)} + \left\langle D_\pi v^j,\ D_{q_i(\theta)} \pi \right\rangle, \tag{30}$$

where $\partial v^j / \partial q_i(\theta)$ holds the prior fixed. Here $D_\pi v^j$ denotes the directional derivative of v^j with respect to the prior: for any perturbation $\Delta\pi$ with $\sum_z \Delta\pi(z) = 0$,

$$\left\langle D_\pi v^j, \Delta\pi \right\rangle \equiv \frac{d}{d\varepsilon} v^j(\boldsymbol{Q}, w^j, \pi + \varepsilon \Delta\pi) \Big|_{\varepsilon=0}.$$

The object $D_{q_i(\theta)}\pi$ is the derivative of the prior with respect to $q_i(\theta)$, defined as follows: for any differentiable test function f on the cognition-state space,

$$\int f(z)\, D_{q_i(\theta)} \pi(dz) \equiv \frac{\partial}{\partial q_i(\theta)} \int f(z)\, \pi(dz; \boldsymbol{Q}).$$

Under consistency and finite Θ,

$$\pi(dz; \boldsymbol{Q}) = \sum_{\theta' \in \Theta} \pi_\Theta(\theta')\, \delta_{z_{\theta'}}(dz), \qquad z_{\theta'} \equiv (\theta', \boldsymbol{q}(\theta')).$$

Thus the support of $\pi(\cdot; \boldsymbol{Q})$ is $Z_\pi \equiv \{z_\theta : \theta \in \Theta\}$. Although the weights $\pi_\Theta(\theta)$ are fixed, the locations z_θ depend on $\boldsymbol{Q}$, so the prior varies with taxes through its support. It follows that for any differentiable f,

$$\int f(z)\, D_{q_i(\theta)} \pi(dz) = \pi_\Theta(\theta) \frac{\partial f}{\partial q_i}(z_\theta). \tag{31}$$

Now let ψ^j be as in Theorem 4. That is, ψ^j is characterized (up to an additive constant) by the property that for any perturbation $\Delta\pi$ with $\sum_z \Delta\pi(z) = 0$,

$$\left\langle D_\pi v^j, \Delta\pi \right\rangle \;=\; \frac{d}{d\varepsilon} v^j(\mathbf{Q}, w^j, \pi + \varepsilon\Delta\pi)\Big|_{\varepsilon=0} \;=\; \sum_{z\in Z_\pi} \psi^j(z)\Delta\pi(z)$$

Using (31) with $f = \psi^j$,

$$\left\langle D_\pi v^j,\, D_{q_i(\theta)}\pi \right\rangle \;=\; \int \psi^j(z)\, D_{q_i(\theta)}\pi(dz) \;=\; \pi_\Theta(\theta)\, \frac{\partial\psi^j}{\partial q_i}(z_\theta).$$

Substituting this into (30) gives the expression used in the main text:

$$\frac{d\tilde{v}^j}{dq_i(\theta)} \;=\; \frac{\partial v^j}{\partial q_i(\theta)} \;+\; \pi_\Theta(\theta)\, \frac{\partial\psi^j}{\partial q_i}(z_\theta).$$

Roy's identity gives $\partial v^j/\partial q_i(\theta) \;=\; -v^j_w\, \pi_\Theta(\theta)\bar{x}^j_i(\theta)$. Substituting the above into the government's first-order condition and applying the generalized Slutsky decomposition as in the proof of Theorem 3 yields equation (19).

B.3 Proof of Proposition ??

To be completed.

References

Allcott, H., Lockwood, B., and Taubinsky, D. (2018). Ramsey Strikes Back: Optimal Commodity Taxes and Redistribution in the Presence of Salience Effects. *AEA Papers and Proceedings*, 108:88–92. 1, 3.2

Angeletos, G.-M. and Sastry, K. A. (2025). Inattentive economies. *Journal of Political Economy*, 133(7):000–000. 1, 3.1, 3.1.2, 3.2, 3.4, 4.2

Boccanfuso, J. and Ferey, A. (2023). Inattention and the Taxation Bias. *Journal of the European Economic Association*, page jvad056. 1

Caplin, A. and Dean, M. (2015). Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review*, 105(7):2183–2203. 1

Caplin, A., Dean, M., and Leahy, J. (2022). Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy. *Journal of Political Economy*, 130(6):1676–1715. Publisher: The University of Chicago Press. 1, 3.1.2

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, 99(4):1145–1177. 1, 4.2

Denti, T. (2023). Unrestricted information acquisition. *Theoretical Economics*, 18(3):1101–1140. 2.2, 2.3

Diamond, P. A. (1975). A many-person Ramsey tax rule. *Journal of Public Economics*, 4(4):335–342. 1, 3.4, 3.4, 3.4

Farhi, E. and Gabaix, X. (2020). Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336. 1, 3.2, 3.3

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124(3):969–1010. 1

Flynn, J. P. and Sastry, K. A. (2023). Strategic mistakes. *Journal of Economic Theory*, 212:105704. 1, 2.3, 3.1, 3.1.1

Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409. 2.2

Geanakoplos, J. and Polemarchakis, H. (1986). Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete. In Heller, W., Starr, R., and Starrett, D., editors, *Essays in Honor of Kenneth Arrow*, volume 3, pages 65–95. Cambridge University Press. [CFDP 764 and CFP 652]. 5

Gerritsen, A. (2016). Optimal taxation when people do not maximize well-being. *Journal of Public Economics*, 144:122–139. 1, 3.2

Goldin, J. (2015). Optimal tax salience. *Journal of Public Economics*, 131:115–123. 1, 3.3

Hébert, B. and La'O, J. (2023). Information acquisition, efficiency, and nonfundamental volatility. *Journal of Political Economy*, 131(10):2666–2723. 1, 3.1.2

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298. 2.3

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational Inattention: A Review. *Journal of Economic Literature*, 61(1):226–273. 1

Mirrless, J. (1975). Optimal commodity taxation in a two-class economy. *Journal of Public Economics*, 4(1):27–33. 3.4

Morris, S. and Yang, M. (2022). Coordination and continuous stochastic choice. *The Review of Economic Studies*, 89(5):2687–2722. 2.2

Mullainathan, S., Schwartzstein, J., and Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. *Annu. Rev. Econ.*, 4(1):511–540. 3.3

Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal*, 37(145):47–61. 3.2

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690. 1, 2.3, 2.3

Sims, C. A. (2010). Rational inattention and monetary economics. In *Handbook of monetary economics*, volume 3, pages 155–181. Elsevier. 1

Stahl, D. O. (1990). Entropy control costs and entropic equilibria. *International Journal of Game Theory*, 19:129–138. 2.2