Diamond-Mirrlees meets Sims:
Optimal Taxation with Rational Inattention

George-Mario Angeletos Matias Bayas-Erazo

March 31, 2026

Abstract

We study optimal commodity taxation when consumers are rationally inattentive. In our framework, consumers may be inattentive to prices and taxes, allocate attention unevenly across decisions, and display behavior resembling sparsity or mental accounting. Our central result is that this need not change optimal tax design. Under suitable conditions, a benevolent Ramsey planner should set taxes exactly as in classical public finance: the key sufficient statistics remain the compensated elasticities of market-level demands, not the cognitive mechanisms that generate them. Away from this benchmark, rational inattention provides a novel rationale for state-dependent taxes: by influencing the uncertainty faced by consumers, such taxes can relax the trade-off between revenue extraction and tax distortion and/or simplify prices and ease cognitive costs.

1 Introduction

How should commodity taxes be designed when consumers are inattentive? A plausible answer is that taxes should correct the distortions created by inattention. Another is that a planner should exploit inattention so as to raise revenue with smaller behavioral responses. Both ideas have played a prominent role in recent work on behavioral public finance. This paper takes a different route. We ask how optimal commodity taxation is altered when inattention is *rational*, as in Sims (2003) and a large follow-up literature (e.g., Maćkowiak et al., 2023; Matějka and McKay, 2015; Caplin and Dean, 2015).

We address this question by introducing a general form of rational inattention in the classic, multi-good, multi-agent Ramsey problem (Diamond and Mirrlees, 1971a,b; Diamond, 1975). The government is benevolent, but it must finance public spending through distortionary taxes. Consumers are rational, but they face cognitive frictions when observing, processing, and responding to prices and taxes. Formally, they choose stochastic consumption rules, or equivalently posteriors about prices and other fundamentals, subject to attention costs. This formulation is flexible enough to encompass noisy information acquisition (Sims, 2003; Maćkowiak et al., 2023), costly control (Flynn and Sastry, 2023), and behaviors reminiscent of sparsity, narrow bracketing, or mental accounting (Caplin et al., 2019; Kőszegi and Matějka, 2020; Lian, 2021; Matějka and Sims, 2011). The central question for us is *normative*: should tax policy attempt to correct these behaviors, exploit them, or ignore them?

Irrelevance. Our main result is an *irrelevance* theorem. Under a set of conditions on attention costs, a benevolent planner should design taxes as if consumers were fully attentive. In that benchmark, the classic sufficient-statistics logic of Ramsey (1927), Diamond and Mirrlees (1971b) and Diamond (1975) survives intact. The planner does not need to identify which part of observed behavior reflects tastes and which part reflects inattention; nor does the planner need to infer the internal cognitive mechanism that generates the observed demand system. For tax design, what matters are the same reduced-form objects as in classical public finance: the price elasticities of the market-level demands. Put simply, a benevolent planner should not care what lies "under the hood" of the observed demand.

This irrelevance result is, in our view, the paper's most practical lesson. A real-world tax authority does not observe preferences, attention choices, or cognitive costs. What it can hope to estimate are demand systems and elasticities. Our result says that, under the appropriate conditions, this is enough. In a pedagogical example with a single taxed good and consumer homogeneity (Section 2), the usual inverse-elasticity formula applies: the optimal tax rate is just Λ/ϵ, where Λ parameterizes the social value of tax revenue and ϵ measures the elasticity of the aggregate demand with respect to the aftertax price. In the richer multi-good, heterogeneous-agent environment (Section 3), the optimal-tax system is still characterized by the same tradeoff among revenue, tax distortion and redistribution as that in Diamond and Mirrlees (1971b) and Diamond (1975). Like preferences, inattention can influence the values of the relevant sufficient statistics, but it does not change the tax formula.

The logic is worth stressing. Rational inattention may generate systematic underreaction in some markets, stronger responses in others, and patterns that resemble coarse thinking or mental accounting. But once the planner internalizes the cognitive costs that make such behaviors optimal, inattention is no more a "mistake" (i.e., a market distortion needing correction); instead, the mistake would be to use taxes to undo those behaviors. What is more, even the positive effects of inattention are ambiguous in general. A common presumption is that inattention dampens demand responses and therefore justifies higher taxes. Our analysis shows that this need not be so. Because attention itself responds to incentives, inattention can make demand more elastic and can thereby justify lower taxes.

In the benchmark identified here, the planner can sidestep all these issues and ambiguities and, instead, proceed directly from observed elasticities to optimal taxes. This benchmark rests on two restrictions on attention costs. The first is *state separability* as in Flynn and Sastry (2023): the cognitive cost of optimizing in one state of nature is independent of what happens in another. The second is *invariance* as in Angeletos and Sastry (2025) and Hébert and La'O (2023): attention costs do not depend on how the underlying primitive uncertainty is described or "packaged" in different endogenous outcomes such as prices. Together, these assumptions yield the sharpest version of our irrelevance result: the Ramsey problem in our inattentive economy translates to a classical Ramsey problem, with fully attentive agents and appropriately modified utilities. Our optimal tax formulas then follow from applying Diamond and Mirrlees (1971b) and Diamond (1975) in this as-if attentive economy.

The assumptions behind our benchmark result are restrictive, but they are neither ad hoc nor empty. In particular, invariance is satisfied by the canonical mutual-information specification used in a large literature, while state separability helps accommodate costly control, or imperfect optimization, even when agents face no uncertainty. Furthermore, our benchmark is instructive because it sets the stage for our paper's second main message: the optimality of state-dependent taxes.

State-dependent taxes. When our benchmark fails, the policy implication is not a blanket shift toward higher or lower taxes. Instead, it is a call for state-dependent taxes—namely of tax rates that fluctuate with realized prices. Two distinct but complementary mechanisms lead to this conclusion.

First, suppose that invariance holds but state separability fails—the case pertinent to mutual-information costs. This does not overturn the irrelevance logic entirely: the planner still does not treat inattention itself as a mistake to be corrected. But it does alter the structure of the optimal tax problem. To understand why, note first that consumers' optimal attention choice naturally depends on the variance of prices, or more generally on the endogenous uncertainty faced by the consumers: the extent of such uncertainty determines the benefit of attention. Note next that state-dependent taxes naturally influence the magnitude of this uncertainty. Putting the two together, we have that state-dependent taxes allow the planner to regulate attention. And while this by itself is a purely positive result, it has the following normative implication: by stimulating more attention, and reducing mistakes in consumption, the planner may better balance revenue extraction against the distortions generated by taxes. In short, when invariance holds but state separability fails, state-dependent taxes improve on the Ramsey tradeoff.

Second, suppose invariance fails. In this case, state-dependent taxes can serve an additional purpose: they can reshape attention costs. When invariance fails, people can not costlessly see through different transformations of the same underlying uncertainty. As a result, laissez-faire prices may be excessively volatile or excessively complex, in the sense that they impose too much cognitive effort on households. A new Pigouvian consideration then appears: optimal taxes should aim at stabilizing or simplifying prices, so as to ease the cognitive burden. The policy goal is different, but the instrument is the same: by regulating the endogenous uncertainty, state-dependent taxes can provide this Pigouvian function in addition to the aforementioned Ramsey function.

Discussion. Our paper has two main lessons. The first is that rational inattention, by itself, does not invalidate the central sufficient-statistics approach of classical public finance. The second is that when attention costs violate the benchmark assumptions, the main new policy implication is that taxes may optimally become state-dependent, either to manage cross-state distortions or to alleviate cognitive burdens. These two lessons fit naturally together. The irrelevance result tells us when the standard Ramsey toolkit remains adequate. The state-dependence result tells us where to look when it does not.

This framing also helps sharpen the contrast with behavioral public finance. In that literature, inattentive or sparse behavior often creates a direct motive for corrective taxation. In our framework, such a motive does not arise merely because observed behavior departs from the frictionless benchmark. The reason is methodological as much as substantive: the planner internalizes the attention costs that rationalize the behavior. To the extent that some prior work finds a corrective motive (Farhi and Gabaix, 2020) that result reflects a different welfare criterion—a paternalistic planner who does not respect the agents' revealed preferences. Under the non-paternalistic discipline imposed by rational inattention, behaviors such as sparsity, mental accounting, or narrow bracketing do not, by themselves, call for correction. Only when the structure of attention costs generates a genuine externality or a nontrivial cross-state interaction does a new rationale for tax design emerge—and this rationale points to state dependence, not to higher or lower taxes.

Related literature. The rational-inattention framework pioneered by Sims (2003) has found a wide range of applications in macroeconomics, finance, and beyond. In the context of individual choice, which is relevant to us, core contributions include Matějka and McKay (2015), Caplin and Dean (2015), and Caplin et al. (2022), which build tight connections between rational inattention, stochastic choice, and revealed preference; and Matějka and Sims (2011), Kőszegi and Matějka (2020), Lian (2021), and Stevens (2015), which show how to place mental accounting and coarse behavior under the rational-inattention umbrella. While we echo the positive insights of these papers, we offer new normative lessons about the design of taxes. Turning from individual choice to general equilibrium, Angeletos and Sastry (2025) show that invariance, together with complete markets, yields an appropriate extension of the first welfare theorem to rationally inattentive economies. We borrow from that paper the insight that invariance is conducive to efficiency, but relax the complete-markets assumption, rule out lump-sum taxation, and build a bridge to classical public finance. This bridge is new—and so are our particular insights about state-dependent taxes.

These insights also distinguish our paper from other rationales for state-dependent taxes found in the literature. The most familiar example comes from the theory of optimal monetary policy (Galí, 2008; Correia et al., 2008): when business cycles are inefficient, due to variable monopoly or other market distortions, a Ramsey planner uses a state-contingent subsidy to offset that variable distortion, or otherwise uses monetary policy to substitute for that tax instrument. In our context, there is no comparable distortion and state-dependent taxes serve a different role: they regulate optimal attention and/or ease the agents' cognitive burden. Another example comes from Angeletos and Pavan (2009): a state-contingent tax helps regulate the use of dispersed private information and thereby the aggregation of such information through endogenous public variables such as asset prices. In our context, this mechanism is moot because consumers try to learn *about* prices as opposed to learning *from* prices, and state-dependent taxes instead serve the functions described earlier.

Last but not least, our paper connects to a growing literature on behavioral public finance (Chetty et al., 2009; Taubinsky and Rees-Jones, 2017; Farhi and Gabaix, 2020). As already discussed, our non-paternalistic stance rules out the particular corrective motive emphasized in some of that literature; and while violations of invariance can introduce a new such motive, quantifying these violations and the resulting state-dependence in optimal taxes is an unexplored territory. Finally, Boccanfuso and Ferey (2023) are close to us methodologically, but study a separate question: how rational inattention may introduce time-inconsistency in optimal policy and in particular a bias towards high taxes.

Layout. The rest of the paper develops our contribution in stages. In Section 2, we illustrate the crux of our irrelevance result in a simple economy with a single taxed good. In Section 3, we study a general multi-good, heterogeneous-agent setting; we establish first the sharpest version of the irrelevance theorem, under invariance and state separability; and we show next how a relaxation of state separability provides a Ramsey rationale for state dependent taxes. In Section 4, we relax invariance and show how this gives a complementary Pigouvian rationale for state dependent taxes. In Section 5, we finally discusses the main takeaways and the connection to behavioral public finance.

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 our irrelevance result, and the role played by endogenous attention in shaping the elasticity of demand, in a transparent way. And while the framework of this section is very stark, it will set the stage for the more general analysis of the subsequent sections.

2.1 Basic environment

The economy is populated by a single type of consumers, who all have the same quasi-linear preferences over two goods, the same income, and the same attention costs. One of the goods is taxed; the other is untaxed and serves as the numeraire. We denote the consumption of the taxed good by x, its pre-tax price (i.e., the producer price) by p, and its aftertax price (i.e., the consumer price) by $q = p+\tau$, where τ is the tax. The consumption of the untaxed numeraire is denoted by y and its price is normalized to 1. The realized utility, gross of attention costs, is $u(x) + y$, where u is a strictly increasing and strictly concave function, with domain $\mathscr{X} \subset \mathbb{R}_+$. The consumer's budget is $qx + y = w$, where w is the common endowment of the numeraire (equivalently, the consumer's wealth).

We fix $w = 1$ and let the aggregate supply of the taxed good be perfectly elastic, which amounts to treating the pre-tax price p as exogenous. This could be justified as follows: the taxed good is produced out of the numeraire good via a linear technology, whose productivity pins down p. As shown in Section 3, these and other simplifying assumptions made in this section are without serious loss of generality. What is crucial for our purposes is that prices are random and that consumers have difficulty tracking, and adjusting to, price changes.

Benchmark with full attention. Suppose momentarily that the consumer can condition x on q perfectly and without incurring any cost. For each realization of q, the consumer thus maximizes $u(x) + y$ subject to $qx + y = w$, or equivalently $u(x) - qx$. In this frictionless benchmark, demand is given by $x = x^{\#}(q) \equiv (u')^{-1}(q)$ and its elasticity is determined by the curvature of u. For example, if $u(x) = \frac{1}{1-1/\eta} x^{1-1/\eta}$, demand is $x^{\#}(q) = q^{-\eta}$ and its elasticity is η. As first shown in Ramsey (1927) and reviewed later, this elasticity in turn determines the optimal tax when consumers are fully attentive.

Adding rational inattention. We depart from the above benchmark by letting the consumer choose a *stochastic* mapping from prices q to consumption x, subject to a cost that depends flexibly on this mapping. The randomness in x conditional on q can then be interpreted interchangeably as noisy information (e.g., Sims, 2003; Matějka and McKay, 2015), noisy cognition (e.g., Woodford, 2020), or imprecise control (e.g., Stahl, 1990; Flynn and Sastry, 2023).

Previous works such as Caplin and Dean (2015) and Fudenberg et al. (2015) have shown how to accommodate such a flexible form of rational inattention in a decision-theoretic context. To adapt that approach to a general-equilibrium context such as ours, one must separate between two types of uncertainty: the uncertainty in prices, which are the relevant fundamentals in the consumer's problem; and the uncertainty in the underlying state of nature, which will determine prices in equilibrium. Denote the exogenous state by $\theta \in \Theta$, where Θ is finite, and let its true, objective, distribution be some $\pi \equiv (\pi_\theta)_{\theta\in\Theta} \in \Delta(\Theta)$, so that π_θ denotes the probability of state θ. To fix ideas, think of θ as an exogenous supply shock. In equilibrium, q will of course be a transformation of θ, but inattentive consumers may or may not have difficulty translating between θ and q. We will expand on this issue in Section 3. For now, we simply let the consumer choose the distribution of x conditional on q subject to a cost.

Specifically, let $\phi \in \Delta(\mathbb{R}_+)$ stand for the consumer's prior about q, denote its support by $\mathcal{Q}_\phi$, and, to ease the exposition, restrict attention to ϕ such that $\mathcal{Q}_\phi$ is finite. Next, let the consumer choose a collection $F = (F(\cdot \mid q))_{q\in\mathcal{Q}_\phi}$, where $F(x \mid q)$ denotes the probability that the consumption of the taxed good is less than or equal to x conditional on the realized price being q; let $\mathcal{F}$ denote the corresponding choice set; and associate each $F \in \mathscr{F}$ with a utility loss $C(F,\phi)$, for some functional C that represents the cost of attention (or cognition, or control). Finally, let the consumption of the numeraire adjust mechanically so as to meet that the consumer's budget (i.e., $y = w - qx$ for all realizations of q and x). Altogether, the individual choice problem can thus be expressed as follows:

$$\max_{F\in\mathscr{F}}\left\{\sum_{q\in\mathscr{Q}_\phi}\int_{\mathscr{X}}\left[u(x)-qx\right]dF(x\mid q)\phi(q)-C(F,\phi)\right\}. \tag{1}$$

In short, the consumer chooses a stochastic demand for x (and hence also for y) so as to maximize expected utility net of attention costs.

The consumer's prior about prices. Note that ϕ, which is a primitive in the consumer's problem, is allowed to enter the costs of attention. This echoes a core idea from the related decision-theoretic literature (e.g., Caplin and Dean, 2015; Caplin et al., 2022; Matějka and McKay, 2015): the cognitive effort an individual must exert to improve upon any given default, to process new information, or to otherwise respond changing circumstances, may naturally depend on what that default is or what the agent knows a priori. Relative to this literature, the novelty here is that ϕ itself is endogenously determined in general equilibrium, as a function of the underlying economy-wide fundamentals and, crucially, of the tax policy. Understanding whether and how the endogeneity of ϕ matters for the *optimal* design of taxes will be an integral part of our analysis.

Examples and interpretations. As already mentioned, our framework can accommodate multiple "flavors" of rational inattention. To illustrate, consider the following example in the spirit of Sims (2003). The consumer conditions her demand on a noisy signal of the price. This signal is $\tilde{q} = q + \sigma\varepsilon$, where ε is noise, drawn from an exogenously specified distribution, and $\sigma > 0$ is chosen by the consumer subject to a cost $K(\sigma)$. In this example, the consumer therefore solves the following problem:

$$\max_{\sigma,x^*(\cdot)}\left\{\mathbb{E}\left[u\left(x^*(q+\sigma\varepsilon)\right)-qx^*(q+\sigma\varepsilon)-K(\sigma)\right]\right\},$$

where $\mathbb{E}$ is integrating over both ε and q. Choosing σ and $x^*(\cdot)$ in this example translates to choosing F in our general formulation, and similarly $K(\sigma)$ translates to $C(F,\phi)$.

Alternatively, suppose that the consumer can observe q perfectly but has trouble implementing the desirable amount of consumption: if, conditional on q, the consumer targets $x = x^*$, the actual consumption is $x = x^* + \sigma\varepsilon$, where ε is an exogenous tremble. Assuming that both x^* and σ can be contingent of q, we can write the consumer problem as follows:

$$\max_{\sigma(\cdot),\, x^*(\cdot)} \left\{\mathbb{E}\left[u\left(x^*(q)+\sigma\varepsilon\right)-q\cdot\left[x^*(q)+\sigma\epsilon\right]-K\left(\sigma(q)\right)\right]\right\}.$$

Choosing F now corresponds to choosing the mean and the variance of x conditional on q, which helps accommodate costly control as, e.g., in Stahl (1990) and Flynn and Sastry (2023).

The optimal taxation result of this section will focus on an economy where rational attention takes the second of the two forms described above. Sections 3 and 4 will allow for an entirely flexible model of rational inattention, nesting both of the above two forms along with many others—including some that help cast sparsity, narrow thinking, and salience under the rational inattention umbrella.

Contrast to behavioral (non-rational) inattention. Consider the forms of sparsity and price misperception considered in Gabaix (2014) and Farhi and Gabaix (2020). There, the consumer consists of two "selves": one choosing the mapping from the actual q to a perceived $\tilde{q}$ subject to a cost, and another choosing consumption conditional on $\tilde{q}$. Whereas the first part mirrors our setting, the second does not. In our context, the consumer recognizes that $\tilde{q} \neq q$ and optimally adjusts the mapping from $\tilde{q}$ to x. In the aforementioned works, instead, consumption is set *as if* the true price was $\tilde{q}$. Put simply, the consumer fails to recognize the imperfection in her perception of prices. These works then focus on a paternalistic planner that, on the one hand, recognizes this imperfection and, on the other hand, disregards the cognitive costs that justified it in the first place. We will clarify these points further when we characterize the optimal taxes in our setting.

2.2 Equilibrium for given taxes

We start by defining equilibrium for given taxes. Since producer prices are exogenously specified in the present setting (but not in the more general setting of Section 3), fixing the taxes amounts to fixing the consumer prices. An equilibrium for given taxes is thus defined by just two requirements: (i) individual optimality; and (ii) consistency of the consumer's prior about prices with their actual distribution. Formally, an equilibrium is a vector $\boldsymbol{Q} \equiv (q_\theta)_{\theta\in\Theta}$, collecting the different realizations of q across the different states of nature, a stochastic choice rule F, and a prior ϕ and such that: (i) F solves problem (1); and (ii) ϕ is such that $\phi(q) = \sum_{\{\theta\in\Theta:\, q_\theta = q\}} \pi_\theta$ for every $q \in \mathbb{R}_+$.

Individual optimality gives F as a function F^* of ϕ. The aggregate demand conditional on q can thus be expressed as $\bar{x}(q,\phi) \equiv \int_{\mathscr{X}} x\, dF^*(x \mid q;\phi)$.[1] Consistency of the prior with the actual price distribution in turn gives ϕ as a function of $\boldsymbol{Q}$. This underscores that, in general, the aggregate consumption in any given state depends not only on the price realized in that state but also on the prices realized in other states, via the consumer's prior. Such interdependence across states vanishes in the special case considered below, but is a necessary implication of mutual-information costs and, more generally, of any cost specification that rationalizes confusion of different price realizations.

2.3 The Ramsey problem

The government is represented by a Ramsey planner, who maximizes social welfare (consumer surplus) plus the shadow value of tax revenue, subject to the restrictions imposed by equilibrium.[2] Let $v(\phi)$ denote consumer surplus, namely the maximum obtained by solving (1); let $\phi^*(\boldsymbol{Q})$ be the equilibrium prior, where, recall, $\boldsymbol{Q} \equiv (q_\theta)_{\theta\in\Theta}$ collects the realizations of the aftertax price across the different states; let $\bar{x}_\theta = \bar{x}(q_\theta, \phi^*(\boldsymbol{Q}))$ and $\tau_\theta = q_\theta - p_\theta$ be the aggregate demand and the tax rate in state θ; and finally let $\lambda > 1$ denote the shadow value of tax revenue. We can then describe the optimal design of taxes, or equivalently the optimal design of $\boldsymbol{Q}$, as the solution of the following Ramsey problem:

$$\max_{\boldsymbol{Q}} \quad v(\phi^*(\boldsymbol{Q})) + \lambda \sum_{\theta\in\Theta} \pi_\theta \left(q_\theta - p_\theta\right) \bar{x}(q_\theta, \phi^*(\boldsymbol{Q})). \tag{2}$$

To interpret this objective, note that the first term captures consumer welfare, net of the cost of attention, whereas the second term measures the social value of tax revenue.[3] Next, note that implicit in this formulation are the two "implementability" constraints faced by the planner: individual behavior is optimal for given ϕ; and ϕ itself is pinned down by $\boldsymbol{Q}$. Finally, note the planner's choice of $\boldsymbol{Q}$ affects, not only the aftertax price of the taxed good in every state, but possibly also the agents' attention costs via the dependence of C on ϕ.

Compared to its "textbook" counterpart, where consumers are fully attentive and make no mistakes, the Ramsey problem described in (2) has three novel features.

1. Individual consumption is subject to noise, reflecting the cost of attention, and both this noise

[1] Throughout, we assume that any noise due to inattention is idiosyncratic and vanishes at the aggregate level.

[2] As usual, the shadow value of tax revenue can be micro-founded by introducing a publicly-provided good and precluding lump-sum taxation.

[3] To understand why we restrict $\lambda > 1$, note that, due to the quasi-linear specification, the marginal value of wealth for the consumer is exactly 1. Had λ been lower than 1, the planner would have found it optimal to rebate any public funds to the consumer. For the optimal taxes to be positive, we must therefore have $\lambda > 1$. This in turn is endogenously guaranteed if we micro-found the social value of tax revenue by adding to the model a publicly provided good, as we do in Section X.

and the attention cost enter the calculation of consumer surplus, as measured by $v(\phi)$.

2. Although the noise washes out across agents, the aggregate demand may still differ markedly from its frictionless counterpart, capturing the *systematic* effect of inattention.

3. Prices may affect aggregate demand and consumer surplus, not only via the familiar substitution and income effects, but by influencing attention choice and attention costs.

At the same time, (2) shows that all these features can be subsumed in appropriate representations of indirect utility and aggregate demand. Leveraging this prism, the next section develops an irrelevance result: under appropriate conditions, inattention matters for optimal taxes exclusively via the price elasticity of aggregate demand. Basically, this is because Roy's identity extends to the presence of inattention, allowing the elasticity of demand to be a sufficient statistic for the effect of taxes on both consumer surplus and tax revenue. A tax authority that accurately estimates this elasticity in the data does not have to discriminate the two underlying forces—preferences vs inattention—nor does it have to adjust the tax formula that the literature has developed for frictionless, fully attentive, economies.

2.4 An irrelevance result

We now establish the aforementioned irrelevance result. A general version of this result and the precise conditions behind it will be spelled out in Section 3. Here, we take a short-cut by imposing the following specification for attention costs:

$$C(F,\phi) = \sum_{q\in\mathcal{Q}_\phi} \phi(q) \int_{\mathcal{X}} K(f(x|q))\,dx, \tag{3}$$

where K is a strictly convex function. This specification, which mirrors our earlier costly-control example, embeds two key assumptions: first, attention costs are "scale free" in the sense that they depend on the prior only through the frequency of different price realizations and not on the exact mapping from the state θ to the price q; and second, attention costs are additively separable across the different realizations of uncertainty. These assumptions will correspond to, respectively, *invariance* and *separability* in Section 3. Here, they yield the following version of our irrelevance result.

Proposition 1. *Suppose attention costs satisfy* (3). *Then, there exists a decreasing function* $\hat{x} : \mathbb{R}_+ \to \mathbb{R}_+$ *such that the aggregate demand of the taxed good is* $\hat{x}(q)$, *where* $q = p + \tau$ *is the aftertax price.*

Furthermore, in every state, the optimal tax satisfies

$$\frac{\tau}{q} = \frac{\Lambda}{\epsilon(q)} \tag{4}$$

where $\Lambda \equiv 1 - \lambda^{-1}$ *is the (rescaled) social value of public funds and* $\epsilon(q) \equiv -\frac{\partial \log \hat{x}(q)}{\partial \log q}$ *is the price elasticity of demand.*

We will prove this result momentarily; first, we explain its economic content and its relation to classic public finance. If we assume away attention costs, demand reduces to $x^{\#}(q) \equiv (u')^{-1}(q)$ and we recover a version of Ramsey (1927): the optimal tax is inversely related to the elasticity of $x^{\#}(\cdot)$, which in turn is determined by the curvature of u. The intuition is familiar: the optimal tax balances the benefit of more tax revenue against the cost of lower consumer surplus; and when u is more concave and demand is thus more elastic, the same incremental increase in the tax crowds out more consumption, implying both a smaller benefit in terms of additional tax revenue and a larger cost in terms of forgone consumer surplus, so the optimal tax is smaller. Proposition 1 shows that this logic carries over here, regardless of how large the agents' attention costs and the resulting mistakes may be: the elasticity of the market-level demand remains a sufficient statistic for optimal taxes.

Proof of Proposition 1. Using (3), we can restate the consumer's problemas follows:

$$v(\phi) = \max_{(f_q)_{q\in\mathcal{Q}_\phi}} \left\{ \sum_{q\in\mathcal{Q}_\phi} \phi(q) \int_{\mathcal{X}} \left[\left(u(x) - qx \right) f_q(x) - K(f_q(x)) \right] dx \right\},$$

where $f_q \equiv f(\cdot|q)$ is a shortcut for the density of x conditional on q . Define now the following two auxiliary functions:

$$\hat{u}(\bar{x}) \equiv \max_f \left\{ \int_{\mathcal{X}} \left(u(x) f(x) - K(f(x)) \right) dx \quad \text{s.t.} \quad \int_{\mathcal{X}} f(x) dx = 1, \quad \int_{\mathcal{X}} x f(x) dx = \bar{x} \right\} \quad \forall \bar{x} \tag{5}$$

$$\hat{v}(q) \equiv \max_{\bar{x}} \left[\hat{u}(\bar{x}) - q\bar{x} \right] \quad \forall q \tag{6}$$

In equilibrium, the consumer's prior about prices is given by $\phi = \phi^*_{\boldsymbol{Q}}$, where $\phi^*_{\boldsymbol{Q}}$ assigns probability $\phi^*_{\boldsymbol{Q}}(q) = \sum_{\{\theta\in\Theta:\, q=q_\theta\}} \pi_\theta$ whenever q is an element of $\boldsymbol{Q}$ (i.e., an equilibrium price) and 0 otherwise. It follows that, in equilibrium,

$$v(\phi^*_{\boldsymbol{Q}}) = \sum_\theta \pi_\theta \hat{v}(q_\theta).$$

This proves that, once restricted to an equilibrium prior about prices, the consumer problem in our economy translates to that in a fictitious economy populated with a representative and fully attentive agent, whose Bernoulli utility is $\hat{u}$ and whose corresponding indirect utility is $\hat{v}$. Intuitively, (5) lets the consumer optimally choose a tremble in her consumption x around a state-contingent target $\bar{x}$ and returns a utility function net of attention costs. It is then *as if* there is a fully attentive agent, just with a modified utility function, which itself subsumes the optimal attention choice.

Although $\hat{u}$ is endogenous to the specification of the attention cost, it is invariant to policy and prices—in the eyes of the planner, $\hat{u}$ ̃is effectively exogenous. It follows that our planner solves the following problem, which is identical to that in the aforementioned fictitious economy:

$$\sum_{\theta\in\Theta} \pi_\theta \max_{q_\theta} \left\{ \hat{v}(q_\theta) + \lambda\left(q_\theta - p_\theta\right)\hat{x}(q_\theta) \right\} \tag{7}$$

with $\hat{x}(q) \equiv \arg\max_x \left[\hat{u}(x) - qx\right]$. The proof is completed by taking the FOCs of this problem, using the envelope condition $\hat{v}'(q) = -\hat{x}(q)$ to arrive at (4), and suppressing the explicit dependence of q and τ on θ.

Inattention and the elasticity of demand. The frictionless benchmark is of course nested in the above argument, just with $x^{\#}(\cdot) \equiv \arg\max_x \left[u(x) - qx\right]$ in place of $\hat{x}(\cdot)$. Because the idiosyncratic noise washes out at the aggregate level, the difference between $\hat{x}(\cdot)$ and $x^{\#}(\cdot)$ isolates the *systematic* effects of inattention. Some literature takes for granted that inattention makes demand less elastic. In an example below, we show that the opposite can be true due to how the optimal attention responds to incentives. This highlights that the relation between inattention and demand elasticity, and thus also that between inattention and optimal taxes, is ambiguous in general. Nonetheless, Proposition 1 establishes that the tax authority does not need to quantify this relation, nor does it need to understand what's going on "under the hood" of the aggregate demand; it suffices to estimate the elasticity of this demand in the standard way and then apply Ramsey's original tax formula.

Multiple goods—a simple case. Proposition 1 readily extends to a multi-good economy in which preferences and attention costs are additively separable, not only across realizations of uncertainty, but also across goods.[4] For each good i, the optimal tax then satisfies

$$\frac{\tau_i}{q_i} = \frac{\Lambda}{\epsilon_i(q_i)}, \tag{8}$$

where $\epsilon_i(q)$, the price elasticity of the aggregate demand for good i, reflects not only the consumers' tastes for good i but also the corresponding attention choice. If consumers are more attentive to certain goods (e.g. groceries or gasoline), and if this translates to the demand for these goods being more sensitive to price changes, then these goods should be taxed less, other things equal. But again, the planner does not have to look "under the hood"; it suffices to measure the elasticity of the demand for each good and apply the usual Ramsey formula. Section 3 will discuss how this insight extends, or changes, when we relax the various forms of additive separability assumed here.

2.5 An example where optimal attention makes demand more elastic

A core theme in the rational-inattention literature is that attention responds to incentives. We now use an example to illustrate how this can justify *lower* taxes than in the frictionless benchmark.

The Bernoulli utility of the single taxed good is $u(x) = \frac{x^{1-1/\eta}-1}{1-1/\eta}$, where $\eta > 0$ is a fixed scalar, $x = \bar{x}e^{\varepsilon}$ is the agent's actual consumption, $\bar{x}$ is its targeted level, $\varepsilon \sim \mathcal{N}(-\frac{1}{2}\nu^{-1}, \nu^{-1})$ is random tremble, and ν is the precision with which this target is attained, or for the level of attention.[5] The consumer chooses both $\bar{x}$ and ν conditional on q, so at to solve the following problem:

$$\max_{\bar{x}(\cdot),\nu(\cdot)} \mathbb{E}\left[u(\bar{x}(q)e^{\varepsilon}) - q\bar{x}(q)e^{\varepsilon} - C(\nu(q))\right], \tag{9}$$

where $C : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing and strictly convex, with $C'(0) = 0$ and $\lim_{\nu\to\infty} C'(\nu) = \infty$. This nests in the proof of Proposition 1 with $\hat{u}(x) \equiv \max_\nu \left[u(R(\nu)x) - C(\nu)\right]$, where $R(\nu) \equiv \left\{\mathbb{E}\left[(e^{\varepsilon})^{1-1/\eta}\right]\right\}^{\frac{1}{1-1/\eta}}$. For the present purposes, though, it is more useful to restate (9) as follows:

$$\max_{\nu}\left[B(\nu,q) - C(\nu)\right] \qquad \text{with} \qquad B(\nu,q) \equiv \max_{\bar{x}}\left(u(R(\nu)\bar{x}) - q\bar{x}\right). \tag{10}$$

The "inner" problem, on the right side, obtains the benefit of attention $B(\nu,q)$ by optimizing over the

[4] This case nests in the setting of Section 3, with $U(\boldsymbol{x}, y) = \sum_{i\in\mathcal{I}} u_i(x_i) + y$ and $C(F,\phi) = \sum_{\boldsymbol{q}\in\mathcal{Q}_\phi} \phi(\boldsymbol{q}) \sum_{i\in\mathcal{I}} \int_{\mathcal{X}} K_i(f_i(x_i|q_i))dx_i$, where $\boldsymbol{x} = (x_i)_{i\in\mathcal{I}}$ and $\mathcal{I}$ is the set of taxed goods.

[5] Note that $\mathbb{E}[x|\bar{x}] = \bar{x}\mathbb{E}[\varepsilon] = \bar{x}$ and $\mathbb{V}[\log x|\bar{x}] = \mathbb{V}[\varepsilon] = \nu^{-1}$, which justify these interpretations. Also note that the present example mirrors the costly-control example of Section 2.1, except that now the tremble is multiplicative and log-normally distributed. This assumption, together with the power form of u, facilitates a closed-form solution.

target $\bar{x}$ for given precision v; the "outer" problem, on the left, trades off this benefit against the cost $C(v)$ to determine the optimal attention v. Let $x^\circ = x^\circ(q, v)$ and $v^\circ = v^\circ(q)$ be the respective solutions of these problems. Then, the effective demand function is $\hat{x}(q) = x^\circ(q, v^\circ(q))$ and its elasticity is

$$\epsilon(q) \equiv -\frac{d \log \hat{x}}{d \log q} = \eta - \frac{\partial \log x^\circ}{\partial v} \frac{d v^\circ}{d \log q}. \tag{11}$$

To understand how this elasticity compares to its frictionless counterpart, we must sign the last term in (11). The following analogy to portfolio choice helps complete this task. As evident from (10). the choice of $\bar{x}$ resembles investment in a portfolio that costs q and yields a risk-adjust return $R(v)$, while the choice of v resembles a diversification strategy that reduces the riskiness of that portfolio at a cost $C(v)$. Under this prism, it is straightforward to see that q has conflicting income and substitution effects. When substitution effects dominate (which is the case for $\eta > 1$), the optimal $\bar{x}$ increases with v because it is optimal to "invest" more when risk is lower, and the optimal v in turn decreases with q because a higher q reduces the optimal investment and hence also the value of reducing that investment's risk. When income effects dominate ($\eta < 1$), *both* of these monotonicities flip. Either way, we end up with $\frac{\partial \log x^\circ}{\partial v} \frac{d v^\circ}{d \log q} < 0$, which proves that the elasticity is *higher* than its frictionless counterpart. By Proposition 1, this comparison extends to the optimal tax. Summing up:

Proposition 2. *In the present example, the elasticity of demand satisfies $\epsilon(q) > \eta$ for all q, and the optimal tax is thus strictly lower than its frictionless counterpart, except for the knife-edge case of $\eta = 1$.*

In Section X we will establish a similar result in an example where inattention takes the form of noisy information acquisition, as opposed to costly control, and we will also nest a preference for sparsity. Together, these examples qualify the presumption in Farhi and Gabaix (2020) and elsewhere that inattention justifies higher taxes by making demand less elastic. At the same time, Proposition 1 and its upcoming generalization allows the planner to completely sidestep the question of how inattention affects demand and, instead, design the optimal taxes *as if* agents were attentive.

3 A General Irrelevance Result

The setting of the previous section was highly restrictive. Does the message of Proposition 1 hinge on the particular assumptions assumptions that setting made about commodity spaces, preferences, and attention costs? In other words, how general is our insight that optimal taxes are determined *as if* agents were attentive, using the same sufficient statistics as those articulated in classic public finance?

We address this question in this section by studying a more general setting. Our starting point is the same class of fully-attentive economies as that studied in Diamond and Mirrlees (1971a,b) and Diamond (1975): we allow for multiple goods, general preferences, and ex-ante heterogeneity. To accommodate a general form of rational inattention to this otherwise-standard public-finance setting, we once again let consumers choose a stochastic mapping from prices (and other fundamentals such as tastes and income) to their consumption, subject to a flexibly specified attention cost. We then go on to define two elementary conditions on attention costs—*state separability* a la Flynn and Sastry (2023) and *invariance* a la Angeletos and Sastry (2025), Caplin et al. (2022) and Hébert and La'O (2023). And we finally characterize optimal taxes depending on whether these conditions hold or not.

The simple economy of Section 2 then nests in a broader class of economies in which both of these conditions hold. In this case, we get the sharpest version of our irrelevance result: regardless of the other model ingredients (preferences, heterogeneity, etc), the optimal taxes are determined, state-by-state, by the exact same formulas as those derived in Diamond and Mirrlees (1971b) and Diamond (1975). Once we relax state separability, consumers may confuse different price realizations—e.g., they may misperceive a high price as a low one. This introduces a certain complexity—demand becomes interdependent across different price realizations—but does not necessarily upset our message: an appropriate version of our irrelevance result continues to apply as long as attention costs are invariant, which in turn is necessarily true with mutual information costs.

To break our invariance result, one must therefore move away from the preferred specification of a large literature so as to relax invariance and open the door to the following possibility: agents' attention costs may depend on how the primitive uncertainty in the economy is endogenously "repackaged" via prices and taxes. We explore this possibility and its policy implications in Section 4. In the present section, we instead focus on establishing the two versions of our irrelevance result—with and without state separability. The upshot is that our irrelevance result can be consistent with price misperception, sparsity, narrow thinking, and mental budgeting, provided that these behaviors are modeled as the product of rational inattention rather than as behavioral biases.

3.1 Set up

There are $N \geq 1$ taxed goods, indexed by $i \in \mathscr{I} \equiv \{1, \ldots, N\}$, along with an untaxed numeraire. There is a measure-one continuum of households split into a finite number of types $h \in \mathscr{H} \equiv \{1, \ldots, H\}$, each with mass μ^h. Different types can have different utilities over these goods and different incomes, as well as different attention costs. We thus accommodate arbitrary ex-ante heterogeneity, not only in terms of tastes and income, but also in terms of the ability to process information, to optimize, or to avoid mistakes. Furthermore, these heterogeneities could be correlated. For example, poor agents could be less attentive, or worse optimizers, than rich agents—or the opposite.[6]

Consumption bundles are denoted by $(\boldsymbol{x}, y)$, where $\boldsymbol{x} = (x_i)_{i\in\mathscr{I}}$ and y are the consumptions of, respectively, the taxed goods and the untaxed good. The latter's price is normalized to 1. The pre-tax (producer) price vector of the taxed goods is denoted by $\boldsymbol{p} \in \mathbb{R}_+^N$, and the corresponding aftertax (consumer) price vector by $\boldsymbol{q} \equiv \boldsymbol{p} + \boldsymbol{\tau} \in \mathbb{R}_+^N$, where $\boldsymbol{\tau} \in \mathbb{R}_+^N$ is the tax vector. The primitive uncertainty behind all these variables is again represented by a "physical" state $\theta \in \Theta$, whose true distribution is denoted by π. The consumers' subjective uncertainty, on other hand, is represented by the "cognitive" state $z = (\theta, \boldsymbol{q})$, whose prior distribution is denoted by ϕ. At the first glance, the distinction between z and θ may appear superfluous, because in equilibrium $\boldsymbol{q}$ will be a deterministic transformation of θ. But with inattentive agents, the distinction can matter, because it may take cognitive effort to adapt behavior to different transformations of the same primitive uncertainty. By distinguishing z from θ, we can flexibly define the individual choice problem and then identify the condition on cognition that renders this distinction immaterial—this condition will be *invariance*, defined in Section 3.2 below.

Turning now to a different issue, note that inattention generates idiosyncratic noise in expenditure. This introduces a familiar complication: how to make sure that inattentive consumers satisfy their budgets. The literature has proposed various ways around this complication. The most standard one is to allow for a quasi-linear good, whose consumption mechanically offsets the noise in the consumption of all other goods. We took this route in Section 2, at the expense of killing the income effects of taxes. Here, we take a different route, which allows for such income effects, in keeping with Diamond-Mirrlees, while also sidestepping the above complication: we assume that the consumer is insured against the noise in her total expenditure. One can think of this in two ways: either literally, with the consumer receiving offsetting lump-sum transfers; or metaphorically, as a proxy for situations where the consumer makes a large multitude of choices, either across many goods or across many periods, and the noise washes out on average.[7] Either way, the essence is that we abstract from

[6]That said, we should also clarify what we do *not* do: we do not study how rational inattention may influence human-capital accumulation or financial investment, thus endogenously affecting the evolution of wealth inequality over time, and we also abstract from ex-post heterogeneity due to uninsurable idiosyncratic shocks.

[7]To see what mean, assume that C is such that the resulting noise is independently distributed across the goods $i \in \{1, ..., N\}$ and take the limit as $N \to \infty$. Then, the consumption of each good can be contaminated with large noise, yet the

uninsurable idiosyncratic variation in the marginal utility of wealth, again in keeping with Diamond-Mirrlees. At the same time, we accommodate arbitrary ex-ante heterogeneity, including in attention costs, as well as incomplete markets vis-a-vis the underlying aggregate uncertainty.[8]

Altogether, we can specify the problem of a type-h consumer as follows:

$$\begin{aligned} \max_{F^h \in \mathscr{F}} \sum_{z \in Z_\phi} \int_{\mathscr{A}} u^h(\boldsymbol{x}, y, \theta) \, dF^h(\boldsymbol{x}, y \mid z) \phi(z) - C^h(F^h, \phi) \\ \text{s.t.} \quad \int_{\mathscr{A}} \left(\boldsymbol{q} \cdot \boldsymbol{x} + y\right) \, dF^h(\boldsymbol{x}, y \mid z) \le w^h_\theta, \quad \forall z \in Z_\phi. \end{aligned} \tag{12}$$

where $\mathscr{A} \subset \mathbb{R}_+^{N+1}$ denotes the commodity space and Z_ϕ denotes the support of ϕ (i.e., the possible states in the consumer's problem). Choice is again stochastic: $F^h(a \mid z)$ is the cumulative density of $a = (\boldsymbol{x}, y)$ conditional on $z = (\theta, \boldsymbol{q})$, and the objective is simply expected utility net of attention costs. Turning to the budget constraints, notice (i) that the consumer faces a separate budget constraint in each z and (ii) the budget in each z is integrating over realizations of the noise in the agents' signal or the tremble in their consumption. These reflect the modeling choices discussed above: there is insurance against idiosyncratic noise, but incomplete markets vis-a-vis variation in prices or other aggregate shocks. Finally, note that, by making different assumptions about the covariation between the consumer's income w^h_θ and the underlying physical state θ (which in turn will drive the equilibrium variation in prices), we can effectively vary the degree of the market incompleteness. This in turn helps clarify the versatility of our upcoming irrelevance result: different assumptions about these primitives, just as about the ex ante heterogeneity, may of course influence the optimal taxes in both attentive and inattentive economies; but as long as the relevant conditions are satisfied, optimal taxes will satisfy the same tax formulas in both types of economies.

3.2 State separability and invariance

We now formalize the two restrictions of attention costs that suffice for the strongest version of our irrelevance result, that is, for the multi-good, multi-agent analogue of Proposition 1. While the formal definitions are somewhat technical, the economic interpretations are straightforward.

noise in total expenditure may vanish.

[8] This distances our analysis from that of Angeletos and Sastry (2025), which instead relies on complete markets across *both* idiosyncratic noise and aggregate shocks in order to prove a version of the First Welfare Theorem.

State separability

Our first condition is imported from Flynn and Sastry (2023) and captures the idea that an agent's cognitive cost of optimizing, or of fine-tuning their demand, in any given state is independent of the cognitive effort exerted in another state. This generalizes the attention cost assumed in Proposition 1 and allows one to recast inattention as a random tremble whose magnitude is determined state-by-state. Formally, we let $a \equiv (\boldsymbol{x}, y) \in \mathscr{A}$ denote a household's action (here, her consumption of both the taxed and the untaxed goods) and define state separability as follows.

Definition 1. The cost of attention, C, is *state separable* if there exists a strictly convex function $K : \mathbb{R}_+ \to \mathbb{R}$ and a weighting function $\alpha : \Theta \times \mathbb{R}_+^N \to \mathbb{R}_{++}$ such that:

$$C(F, \phi) = \sum_{z \in Z_\phi} \alpha(z)\phi(z) \int_{\mathscr{A}} K(f(a \mid z))\, da, \tag{13}$$

for any F with density f.[9]

Intuitively, state separability requires the cost of controlling mistakes in state z (e.g., when the price is high) be independent of what happens in state $z' \neq z$ (e.g., when the price is low). This property therefore reflects a "local" view of inattention, where cognitive resources can be allocated independently across different economic conditions. It is as if the agent maintains separate mental accounts for different situations, effectively facing separate attention allocation problems across different states. Although this is of course restrictive, it can capture a wide range of decision frictions, including state-contingent attention costs, ex-post optimization errors, ex-ante planning frictions, and endogenous consideration sets (Flynn and Sastry, 2023). This is achieved by changing the weighting function α and/or the kernel K.[10] Finally, note that that, although this property amounts to letting cognition be independent across realizations of uncertainty, cognition can still be interdependent across goods or consumption choices; this in turn can change the effective substitutability or complementarity of different goods, and generate behaviors akin to mental accounting (Lian, 2021).

[9] To simplify the exposition, and without any loss of generality, we henceforth assume that F admits a density. Alternatively, one could let F have a finite support $A(z)$ for each z, reinterpret $f(a \mid z)$ as the probability of a conditional on z, replace $\int_{\mathscr{A}} K(f(a \mid z))\, da$ in equation (13) with $\sum_{a \in A(z)} K(f(a \mid z))$, and accordingly adjust the rest of the mathematical arguments in the Appendix, without changing any of the essence.

[10] Furthermore, note that, even if the marginal cost of attention is the same across states, which is the case if $\alpha(z) = 1$ for all z, this does not mean that the optimal attention won't vary with the state z. As we already illustrated in the previous section, different states (e.g., different price realizations) may naturally influence the marginal benefit of attention and thereby its optimal level.

Notwithstanding all these points, state separability has an important limitation: it rules out rational confusion of different states and thus also price misperceptions. Intuitively, this property implies that the agent cannot strictly economize attention costs by correlating mistakes across states, so it is optimal to learn *perfectly* the state and then simply add a tremble to demand. For optimal attention to confound different states, a violation of separability is necessary.

Invariance

Our second condition—*invariance*—imposes that attention costs do not change if we fix the underlying statistical uncertainty and change the way it is described. To define this precisely, we must first introduce two auxiliary concepts—*transformations* and *sufficiency*. These concepts help formalize the idea that the same stochastic choice problem can be described in multiple ways; invariance then requires that attention costs remain unchanged across such equivalent descriptions.[11]

Definition 2 (Transformation)**.** Consider two pairs (F,ϕ) and $(\tilde{F},\tilde{\phi})$, where F and $\tilde{F}$ denote different stochastic choices, while ϕ and $\tilde{\phi}$ denote different priors about z. Next, consider a mapping $g : \Theta \times \mathbb{R}^N_+ \to \Theta \times \mathbb{R}^N_+$. We say that $(\tilde{F},\tilde{\phi})$ is the *transformation* of (F,ϕ) under g if:

$$\tilde{\phi}(z) = \sum_{z' \in Z_\phi} \phi(z')\mathbb{1}_z\{g(z')\}, \quad \forall z, \tag{14}$$

$$\tilde{f}(a \mid z) = \frac{\sum_{z' \in Z_\phi} f(a \mid z')\phi(z')\mathbb{1}_z\{g(z')\}}{\tilde{\phi}(z)}, \quad \forall a,\ z \text{ such that } \tilde{\phi}(z) > 0. \tag{15}$$

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

A transformation relabels or regroups states of the world via the mapping g. For example, one could consider changing the words used to describe the underlying states (say, from "low" and "high" to "bad" and "good"), without changing their probabilities. Alternatively, one may combine finely detailed states into coarser categories (say, group "1,2,3" to "low" and "4,5" to high). The first part of Definition 2 describes the distribution of this new random variable implied by the original one. The second part specifies how the new stochastic choice plan—the distribution of demand conditional on the realization of this new variable—is constructed from the original one. Sufficiency then means

[11] All these concepts can be defined for more abstract settings, along the lines of Angeletos and Sastry (2025) and Hébert and La'O (2023). Here, we adapt them to our setting.

that the transformed plan generates the same conditional distributions of actions as the original plan, possibly under a different labeling or aggregation of states. In particular, the transformed state $g(z)$ is "sufficient" for z in the sense that it carries all the information from z that is relevant for replicating the agent's original stochastic choice. Therefore, while the transformation g may change how states are labelled or packaged together, sufficiency ensures that this does not really change behavior.

Now ask: if we change how the underlying uncertainty is described, but adjust the agent's behavior so as to replicate the original stochastic choice, should we expect their attention costs to remain unchanged? *Invariance* is defined so that the answer to this question is an unequivocal yes.

Definition 4 (Invariance)**.** Let $G \equiv \{g : \Theta \times \mathbb{R}_+^N \to \Theta \times \mathbb{R}_+^N\}$ and consider any two pairs (F,ϕ) and $(\tilde{F},\tilde{\phi})$ such that: $(\tilde{F},\tilde{\phi})$ is a transformation of (F,ϕ) under some $g \in G$; and $\tilde{\phi}$ is sufficient for ϕ with respect to F. The cost of attention is *invariant* if and only if $C(F,\phi) = C(\tilde{F},\tilde{\phi})$, for any such two pairs.

To sum up, when attention costs are invariant, they do not depend on the scale or "alphabet" used to describe the underlying uncertainty. And since equilibrium prices effectively translate the same uncertainty from one alphabet (θ) to another ($\boldsymbol{q}$), invariance means that agents can "see through" this translation and pay no more or less cost for learning about prices (i.e., for correlating their choice with the realized prices) than for learning directly about the physical state of nature.[12]

This property will be the key to the most general version of our irrelevance result. Also note that this property is necessarily satisfied by mutual information costs, because these costs depend only on the densities of actions and states.[13] A concrete counterexample, on the other hand, is one where attention costs depend, not only the mutual information between $a = (\boldsymbol{x}, y)$ and $z = (\theta, \boldsymbol{q})$, but also on the variance of z. As discussed in Section 4, this could capture the idea that agents face a harder cognitive problem when they live in a more uncertain, or more complex, market environment.

3.3 The irrelevance result—with state separability

We now provide the sharpest version of our irrelevance result, which assumes both invariance and state separability.

[12] In fact, as this discussion suggests, we could relax our notion of invariance to apply only to a strict subset of G, given by $G^* \equiv \{g : \Theta \times \mathbb{R}_+^N \to \Theta \times \mathbb{R}_+^N$ such that $g(\theta) = (\theta, h(\theta))$ for all θ and some function $h : \Theta \to \mathbb{R}_+^N\}$. That is, it is sufficient to have invariance over transformations of the price mapping.

[13] See Angeletos and Sastry (2025), Caplin et al. (2022), and Hébert and La'O (2023) for discussions of how the concept of invariance used here connects to mutual information costs, as well as to related notions of invariance in the literature.

We follow a similar procedure as in Section 2. We start by mapping our economy to another, fictitious economy, populated by attentive agents with appropriately modified utilities. There are now H types of such as-if consumers, each one solving the following problem:

$$\max_{\boldsymbol{X}^h, \boldsymbol{Y}^h} \left\{ U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi) \text{ s.t. } \boldsymbol{Q} \cdot \boldsymbol{X}^h + \boldsymbol{Y}^h \leq \boldsymbol{W}^h \right\}, \tag{16}$$

where $\boldsymbol{X}^h \equiv \left(\boldsymbol{x}_\theta^h\right)_{\theta\in\Theta}$ and $\boldsymbol{Y}^h \equiv \left(y_\theta^h\right)_{\theta\in\Theta}$ collects the consumptions of the taxed goods and the untaxed numeraire in the various states, $\boldsymbol{Q} \equiv \left(\boldsymbol{q}_\theta\right)_{\theta\in\Theta}$ and $\boldsymbol{W}^h \equiv (w_\theta^h)_{\theta\in\Theta}$ collect the corresponding prices and incomes, $U^h(\cdot, \phi)$ is a utility function describing the preferences of the fictitious attentive agent over $\boldsymbol{X}^h$ given a prior $\phi \in \Phi$, and Φ denotes the class of priors over cognition states whose support is spanned by θ.[14] Differently from Section 2, the fictitious agent here consumes multiple goods, does not have quasi-linear preferences over a "residual" good, and faces multiple budget constraints, one for each state θ. Similarly to Section 2, U^h embeds, not only the true, primitive preferences of group h in our economy, but also their costs of attention. Letting $\phi^*(\boldsymbol{Q})$ denote the unique prior consistent with $\boldsymbol{Q}$, we thus have the following equivalence between the two economies.

Lemma 1. *For each h, there exists an as-if utility function U^h such that, in equilibrium, the average demand of type-h consumers is $\boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q}))$ and their expected utility, net of attention costs, is $V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q}))$, with the functions $\boldsymbol{X}^h$ and V^h defined as the* arg max *and the* max *of* (16).

Given this result, the Ramsey problem of our general model can be expressed as follows:

$$\max_{(\boldsymbol{Q},\phi)} \left\{ \sum_{h\in\mathcal{H}} \mu^h \beta^h V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi) + \sum_{\theta\in\Theta} \tilde{\lambda}_\theta \left(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\right) \cdot \sum_{h\in\mathcal{H}} \mu^h \boldsymbol{X}^h(\theta, \boldsymbol{Q}, \boldsymbol{W}^h, \phi) \quad \text{s.t.} \quad \phi = \phi^*(\boldsymbol{Q}) \right\},$$

where $\beta^h > 0$ and $\tilde{\lambda}_\theta > 0$ represent, respectively, the Pareto weight of type-h agents and the shadow value of tax revenue in state θ. Compared to Section 2, the planner recognizes that taxes on different goods may have different incidence across different types of agents, and she furthermore faces a more complex demand structure: there can be rich interdependences, not only across goods, but also across states of nature, due to rational confusion. Nevertheless, the following two lessons carry over from Section 2. First, if attention costs are invariant, we can drop the prior ϕ from the corresponding U^h: agents can "see through" different transformations of the underlying uncertainty. Second, if attention costs are also separable, U^h admits an expected utility representation: there exists a (possibly

[14] The exact definitions of Φ and U^h are deferred to Appendix **??**.

state-dependent) Bernoulli utility $\hat{u}^h : \Theta \times \mathscr{A} \to \mathbb{R}$ such that $U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi) = \sum_{\theta\in\Theta} \pi_\theta \hat{u}^h(\theta, \boldsymbol{x}_\theta^h, y_\theta^h)$, for all $\boldsymbol{X}^h \equiv \left(\boldsymbol{x}_\theta^h\right)_{\theta\in\Theta}$, all $\boldsymbol{Y}^h \equiv \left(y_\theta^h\right)_{\theta\in\Theta}$ and all $\phi \in \Phi$. By the same token, it is *as if* type-h agents solve a separate problem in each state, given by

$$\max_{\boldsymbol{x}_\theta^h, y_\theta^h} \left\{ \hat{u}^h(\theta, \boldsymbol{x}_\theta^h, y_\theta^h) \text{ s.t. } \boldsymbol{q}_\theta \cdot \boldsymbol{x}_\theta^h + y_\theta^h \le w_\theta^h \right\}. \tag{17}$$

We can therefore simplify the objects in the planner's problem as follows:

Lemma 2. *Suppose C^h is both invariant and state separable. Then, there exists a Bernouli utility $\hat{u}^h$: $\Theta \times \mathscr{A} \to \mathbb{R}$ such that, for all $(\boldsymbol{Q}, \boldsymbol{W}^h)$ and for $\phi = \phi^*(\boldsymbol{Q})$,*

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi) = \sum_{\theta\in\Theta} \pi_\theta \hat{v}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h) \qquad \textit{and} \qquad \boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q})) = \left(\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)\right)_{\theta\in\Theta}$$

where $\hat{v}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)$ and $\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)$ are, respectively, the max *and the* arg max *of problem* (17).

Provided that this holds for all groups, our planner solves, in effect, a separate policy problem for each θ. In particular, the optimal taxes in state θ—equivalently, the optimal aftertax prices $\boldsymbol{q}_\theta$—solve the following problem:

$$\max_{\boldsymbol{q}_\theta} \left\{ \sum_{h\in\mathscr{H}} \mu^h \beta^h \hat{v}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h) + \lambda_\theta \left(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\right) \cdot \sum_{h\in\mathscr{H}} \mu^h \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h) \right\},$$

with $\lambda_\theta \equiv \tilde{\lambda}_\theta / \pi_\theta$. For each θ, this is mathematically identical to the Ramsey problems studied in Diamond and Mirrlees (1971b) and Diamond (1975), so we have basically proved our irrelevance result.

To precisely state this result, we only need some additional notation. Following Diamond (1975), we define the marginal *social* value of income for group h in state θ as follows:

$$\gamma_\theta^h \equiv \beta^h \partial_w v_\theta^h + \lambda_\theta \boldsymbol{\tau}_\theta \cdot \partial_w \bar{\boldsymbol{x}}_\theta^h, \tag{18}$$

where $\partial_w v_\theta^h \equiv \frac{\partial \hat{v}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)}{\partial w}$ is the marginal *private value* of income in state θ and $\partial_w \bar{\boldsymbol{x}}_\theta^h \equiv \frac{\partial \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)}{\partial w}$ is the vector of income effects on demand. This definition accounts for how a change in the disposable income of group h affects social welfare in two ways: by changing the utility of those agents (the first term); and by changing their spending on taxed goods and thus tax revenue (the second term, known as a "fiscal externality"). Next, we define the Slutsky matrix for group h in state θ by

$$\boldsymbol{S}_\theta^h \equiv \partial_{\boldsymbol{q}} \bar{\boldsymbol{x}}_\theta^h + (\partial_w \bar{\boldsymbol{x}}_\theta^h)(\bar{\boldsymbol{x}}_\theta^h)^\top \tag{19}$$

where $\partial_{\boldsymbol{q}} \bar{\boldsymbol{x}}_\theta^h \equiv \left(\frac{\partial \hat{x}_i^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)}{\partial q_j}\right)_{i,j \in \mathscr{I}}$ is the matrix of price effects, $\partial_w \bar{\boldsymbol{x}}_\theta^h$ is the vector of income effects, and $\bar{\boldsymbol{x}}_\theta^h = \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h)$ are the demanded quantities.[15] We can then state the following irrelevance result.

Theorem 1. *Let attention costs be both invariant and state-separable. The optimal taxes then satisfy*

$$\sum_{h \in \mathscr{H}} \mu^h \left[\left(\lambda_\theta - \gamma_\theta^h\right) \bar{\boldsymbol{x}}_\theta^h + \lambda_\theta \boldsymbol{S}_\theta^h \boldsymbol{\tau}_\theta \right] = \boldsymbol{0} \qquad \forall \theta. \tag{20}$$

Equation (20) mirrors the tax formula developed in Diamond (1975). Just as in that paper, a marginal increase in $\boldsymbol{\tau}_\theta$ affects social welfare through three standard channels. First, it increases the tax revenue collected from type-h agents by $\bar{\boldsymbol{x}}_\theta^h$, for all h; the social value of the additional revenue is $\lambda_\theta \sum_{h=1}^H \mu^h \bar{\boldsymbol{x}}_\theta^h$. Second, it makes type-$h$ agents poorer by $\bar{\boldsymbol{x}}_\theta^h$, again for all h; this lowers social welfare by $\sum_{h=1}^H \mu^h \gamma_\theta^h \bar{\boldsymbol{x}}_\theta^h$. Finally, the Slutsky terms capture the distortionary effects of taxation when agents substitute away from taxed commodities. At the optimum, the planner sets taxes so that the sum of these effects cancels out. Crucially, this logic and the associated tax formula are *exactly* the same whether agents are attentive or inattentive—the agents' attention costs and their optimal attention choices have been subsumed in the reduced-form demands.

Theorem 1 thus extends the insight of Proposition 1 to a general, multi-good, multi-agent setting. To further illustrate this connection, abstract from redistributive motives and let producer prices be the only source of uncertainty. This means $\gamma_\theta^h = 1$ for all h,θ; $\lambda_\theta = \lambda > 1$ and $w_\theta^h = w^h > 0$ for all θ; and, with abuse of notation, $\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w_\theta^h) = \hat{\boldsymbol{x}}^h(\boldsymbol{q}_\theta)$ for all θ, that is, demand varies *only* because of price variation. Next, let $\hat{\boldsymbol{x}}(\boldsymbol{q}) \equiv \sum_{h \in \mathscr{H}} \mu^h \hat{\boldsymbol{x}}^h(\boldsymbol{q})$ and $\boldsymbol{S}(\boldsymbol{q})$ be, respectively, the aggregate demand function and the corresponding Slutsky matrix; suppose that $\boldsymbol{S}(\boldsymbol{q})$ is invertible; and finally let $\Lambda \equiv 1 - \lambda^{-1}$ be the rescaled social value of funds. We can then specialize Theorem 1 to a much simpler tax formula.

Corollary 1. *Consider the class of economies described above, where attention costs are invariant and state-separable, distributional concerns are absent, and all uncertainty is about producer prices. The*

[15] Note that both γ_θ^h and $\boldsymbol{S}_\theta^h$ depend on the underlying allocation and thus on the optimal taxes. We suppress this dependence, which is inherited from fully-attentive economies, to ease the notation. Also note that, by implication of Lemma 2, the Slutsky matrix is symmetric, despite the introduction of inattention. This symmetry, however, can break once attention costs are non-separable or non-invariant.

optimal taxes then satisfy

$$\boldsymbol{\tau} = -\Lambda(\boldsymbol{S}(\boldsymbol{q}))^{-1}\hat{\boldsymbol{x}}(\boldsymbol{q}), \tag{21}$$

Compared to Proposition 1, the key changes here are the use of compensated elasticities, to net out income effects, and the accommodation of a general, non-diagonal, Slutsky matrix, to account for substitutability or complementarity across goods. These adjustments are familiar from classic public finance, but there is now a twist: substitutability or complementarity may now reflect, not only the agents' tastes, but also their attention choices. Borrowing insight from Kőszegi and Matějka (2020) and Lian (2021), we next relate this basic point to "narrow bracketing" and "mental accounting".

Inattention, mental accounting, and optimal taxation

Revisit the example of Section 2.5, allow for two taxed goods, and assume that utility is additively separable—which means that the demands and so the expenditures of the two goods are independent in the frictionless benchmark. Following the same steps as before, we can reduce the inattentive consumer's problem to the following:

$$\max_{(\nu_1,\nu_2,\rho)} \left[B_1(\nu_1,q_1) + B_2(\nu_2,q_2) - C(\nu_1,\nu_1)\right],$$

where ν_i is the attention, or the precision, allocated to good i, $B_i(\cdot)$ is the corresponding benefit, ρ is the correlation of the trembles across the two choices, and $C(\cdot)$ is the attention cost.[16] Note that the benefits of attention are separable across ν_1 and ν_2, because of the separability in the underlying utility, but the costs need not be separable. For example, suppose $C(\nu_1,\nu_1) = (\nu_1+\nu_1)^k$ for some $k > 1$; this could capture a situation where it is costly to increase overall cognitive capacity (precision) but costless to allocate it across decisions. As explained in Section (2.5), a higher q_1 reduces the optimal ν_1; but as this now frees cognitive capacity, the optimal ν_2 increases, which in turn boosts the demand for good 2. (And symmetrically, a higher q_2 boosts the demand for good 1.) Optimal attention therefore turns the two goods from independent to substitutes. Furthermore, because $\eta_i > 1$ guarantees that the expenditure on good 1 falls with q_i, and because now this is accompanied with higher expenditure on good $j \neq i$, the consumer may appear to be reallocating a fixed budget to the goods.

[16] To fill in the details, let utility be $u_1(x_1) + u_2(x_2) + y$, where $u_i(x) = \frac{x^{1-1/\eta_i}-1}{1-1/\eta_i}$ and $\eta_i > 1$ (which means that substitution effects dominate) for both i. Next, for each i, let the realized consumption of good i be $x_i = \bar{x}_i e^{\varepsilon_i}$, where $\bar{x}_i$ is the targeted level, $\varepsilon_i \sim \mathcal{N}(-\frac{1}{2}\nu_i^{-1}, \nu_i^{-1})$ is the tremble, and $\nu_i \equiv 1/\mathbb{V}[\varepsilon_i]$ is the corresponding precision. Then, $B_i(\nu_1,q_1) \equiv \max_{\bar{x}}\left(u_i(R_i(\nu)\bar{x}) - q\bar{x}\right)$, where $R_i(\nu) \equiv u_i^{-1}(\mathbb{E}[u_i(e^{\varepsilon})])$. Also, for simplicity, we assume that

Using richer examples, Kőszegi and Matějka (2020) and Lian (2021) argue that rational inattention can help accommodate behaviors akin to "narrow bracketing" and "mental accounting". Here, we complement their positive results with a normative lesson: if these non-standard behaviors are the product of rational inattention, they may not require any correction by the planner and they may not affect the design of optimal taxes. We reinforce this message in the next section by accommodating noisy learning and price misperceptions, further bridging our analysis to those papers.

3.4 The irrelevance result—without state separability

We now relax state separability, while maintaining invariance. In this case, which nests mutual information costs, agents confuse one state for another—e.g., the consumer may misperceive a high tax for a low one. This misperception is, however, rational in a dual sense: first, it is the byproduct of optimal learning; and second, when the agents chooses consumption, she recognizes that her perception of aftertax prices could be wrong. We now show how this allows an appropriate version of our irrelevance result to continue to apply.

As long as attention costs are invariant, the following property survives from the argument behind Theorem 1: the indirect utility and demand functions of the fictitious attentive agents representing each group h, as defined in Section 3, do not depend on the prior ϕ. The only material difference is that the optimal attention and thus the optimal demands in state θ depend on the entire vector $(\boldsymbol{Q}, \boldsymbol{W}^h)$, as opposed to merely the same-state realizations $(\boldsymbol{q}_\theta, w^h_\theta)$. Intuitively, because attention costs are not state-separable, the optimal attention and by extension the optimal consumption in any state depends on the incentives faced in other states.

To accommodate this complication, we introduce the following cross-state derivatives: we let $\partial_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \left(\frac{\partial \bar{x}^h_i(\theta', \boldsymbol{Q}, \boldsymbol{W}^h)}{\partial q_{\theta,j}}\right)_{i,j\in\mathscr{I}}$ denote the response of demand in state θ' to prices in state θ, and we similarly let $\partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \frac{\partial \bar{\boldsymbol{x}}^h(\theta', \boldsymbol{Q}, \boldsymbol{W}^h)}{\partial w_\theta}$ denote the corresponding cross-state income effect. We then define the cross-state Slutsky matrix for type h as follows:

$$\boldsymbol{S}^h_{\theta,\theta'} \equiv \partial_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_{\theta'} + \left(\partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'}\right)\left(\bar{\boldsymbol{x}}^h_\theta\right)^\top.$$

This of course nests the within-state Slutsky matrix encountered in Section 3 when $\theta' = \theta$; but whereas before we had effectively restricted $\boldsymbol{S}^h_{\theta,\theta'} = \boldsymbol{0}$ for all $\theta' \neq \theta$, we now allow $\boldsymbol{S}^h_{\theta,\theta'} \neq \boldsymbol{0}$. We similarly adjust the marginal social value of income, to account for nay cross-state fiscal externalities:

$$\gamma_\theta^h \equiv \beta^h \partial_{w_\theta} V^h + \sum_{\theta' \in \Theta} \lambda_{\theta'} \boldsymbol{\tau}_{\theta'} \cdot \partial_{w_\theta} \bar{\boldsymbol{x}}_{\theta'}^h.$$

With these adjustments, we obtain the following generalization of Theorem 1.

Theorem 2. *Let attention costs be invariant (but not state-separable). The optimal taxes then satisfy*

$$\sum_{h \in \mathcal{H}} \mu^h \left[\left(\lambda_\theta - \gamma_\theta^h\right) \bar{\boldsymbol{x}}_\theta^h + \sum_{\theta' \in \Theta} \lambda_{\theta'} \left(\boldsymbol{S}_{\theta,\theta'}^h\right)^\top \boldsymbol{\tau}_{\theta'} \right] = \boldsymbol{0} \qquad \forall \theta. \tag{22}$$

Compared to Theorem 1, the only material difference here is the inclusion of cross-state price elasticities, as captured in $\boldsymbol{S}_{\theta,\theta'}^h$ for $\theta' \neq \theta$. While these terms are generally non-zero, the high-level economics remains unchanged: the planner still trades off revenue gains against deadweight losses. This trade off is no more separable across states, but this is a quantitative, not a conceptual, difference. Indeed, as evident from our fictitious-economy representation, relaxing state-separable attention costs in our context is equivalent to relaxing expected utility in the traditional context.

In short, the present case—which nests mutual-information costs—preserves the essence of our earlier irrelevance result, while at the same time allowing inattention to create rich patterns of substitutability or complementarity, not only across goods, but also across realizations of uncertainty. As anticipated in the previous section, this further aligns our normative analysis with the positive contributions of Kőszegi and Matějka (2020) and Lian (2021) and lets our result to speak to "narrow bracketing" and "mental accounting", as well as to price misperceptions. The upshot is that, if inattention is rational and if invariance holds, the planner should *not* try to correct these behaviors.

Notwithstanding these points, non-separable costs raise an empirical challenge: how to quantify the resulting cross-state effects on demand. To make this issue more tangible, consider again the special case studied in Corollary 1, that is, abstract from distributional concerns and equate "states" with "price realizations". The statement "demand is interdependent across states" then translates to the following simpler statement: demand depends not only on the realized price but also on the distribution of all other possible price realizations. Suppose now that this dependence can be summarized via a (non-random) vector $\boldsymbol{\sigma} = (\sigma_k)_{k=1}^K \in \mathbb{R}^K$, capturing some moments of that distribution and defined flexibly as $\sigma_k \equiv \mathbb{E}[m_k(\boldsymbol{q})] = \sum_\theta \pi_\theta m_k(\boldsymbol{q}_\theta)$ for some function $m_k : \mathbb{R}^N \to \mathbb{R}$, for $k \in \{0, \ldots, K\}$ and $K \geq 1$. Re-writing aggregate demand as $\hat{\boldsymbol{x}}(\boldsymbol{q}, \boldsymbol{\sigma})$ and the corresponding Slutsky as $\boldsymbol{S}(\boldsymbol{q}, \boldsymbol{\sigma})$, we can then recast Theorem 2 as follows.

Corollary 2. *Consider the class of economies described above, where attention costs are invariant, distributional concerns are absent, all uncertainty is about producer prices, and the state-interdependence in demand is summarized in the moments* $\boldsymbol{\sigma}$*. The optimal taxes then satisfy*

$$\boldsymbol{S}(\boldsymbol{q},\boldsymbol{\sigma})^{\top}\boldsymbol{\tau}+\sum_k \mathbb{E}[\partial_{\sigma_k}\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma})\boldsymbol{\tau}]\partial_{\boldsymbol{q}}m_k(\boldsymbol{q})=-\Lambda\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma}). \tag{23}$$

This qualifies the irrelevance result of Theorem 2 as follows: although the tax formula (22) coincides with that of a fully-attentive economy, this formula looks non-standard once recast in the form of (23): it contains an unfamiliar term (the second term seen above), capturing how taxes affect $\boldsymbol{\sigma}$, which in turn affects the *expected* tax revenue. Conceptually, this new term is *not* different from the first term in (23): both terms capture how a marginal change in taxes affects tax revenue. Still, this term poses an empirical challenge: to implement (23), the planner would need to estimate, not only the usual compensated elasticities, but also how optimal attention and thus demand vary with $\boldsymbol{\sigma}$, the variance or other moments of the aftertax prices.

Given this empirical challenge, a real-world tax authority may ignore the new term and instead take guidance from the simpler, totally standard, tax formula in Corollary 1. In this sense, the strongest version of our irrelevance result remains our main practical takeaway. That said, we next use a "textbook" example, augmented with mutual information costs, to shed further light on why rational inattention invites state-dependent taxes.

3.5 An example with mutual-information costs

We revisit the example of Section 2.5, which features a single taxed good along with a untaxed quasilinear numeraire. As in that section, the net utility of consuming x units of the tax good is is $u(x)-qx$, with $u(x)\equiv\frac{1}{1-1/\eta}x^{1-1/\eta}$ and $\eta>0$. Unlike that section, attention takes the form of the choice of a signal about q, subject to a cost that depends on the mutual information between that signal and q.

To facilitate a closed-form solution, we make four assumptions. First, we let the exogenous, pretax, producer price be log-normally distributed: $p=\theta$ and $\log\theta\sim\mathcal{N}(0,v_\theta)$, for some $v_\theta>0$. Second, we restrict attention to a tax such that the aftertax price q is a log-linear transformation of p:

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

where χ and φ parameterize, respectively, the mean value and the state-dependence, or cyclicality, of the tax rate. Third, we let the signal be $\omega = \log q + \varepsilon$, where ε is a Normal noise, and we model attention as the choice of the variance of this noise or equivalently of the value of $\rho \equiv \frac{\mathbb{V}[\log q]}{\mathbb{V}[\epsilon]+\mathbb{V}[\log q]} \in [0,1]$; this is the same as choosing the mutual information of ω and q. Finally, we let the agent's overall payoff be

$$\mathscr{U} = \mathbb{E}\left[u(x(\omega)) - qx(\omega)\right] \exp\left\{-(1-\tfrac{1}{\eta})K(\rho)\right\}, \tag{24}$$

where $K(\rho)$ is a strictly increasing and convex function, capturing the costs of attention.[17]

In this example, the optimal individual consumption is given by

$$x(\omega) = \left(\mathbb{E}[q|\omega]\right)^{-\eta} = \exp\left\{-\eta\left(\rho\omega + (1-\rho)\log\mathbb{E}[q]\right)\right\}. \tag{25}$$

Had ρ being exogenous, the price elasticity of the *aggregate* demand would have equaled $\eta\rho$, and lower attention would have thus translated to a lower elasticity and a higher tax (here, higher χ), consistent with the conventional view. But ρ is endogenous here, and the question of interest is how this enters for the optimal state-dependence of the tax (φ). To address this question, we first study how the agents' attention choice depends on the stochastic properties of the aftertax prices.

Proposition 3. *Consider the example with mutual-information costs described above. The optimal ρ increases with* $\mathbb{V}(\log q)$: *consumers pay more attention when aftertax prices are more volatile.*

Intuitively, choosing a ρ amounts to choosing the correlation between consumption and prices. For given $\rho < 1$, more volatile prices translate to a larger consumption mistakes, lower expected utility, and a higher benefit of paying more attention, which explains the above result. Note then that $\mathbb{V}(\log q) = \varphi^2 v_\theta$. Hence, although the *level* of taxation, as parameterized by χ, does not influence ρ, the degree of *state-dependence* in taxes does: by varying φ, the planner can regulate the agents' choice of ρ. But is it optimal to do so? And if yes, in what direction? The next result answers this question.

Proposition 4. *Consider the example with mutual-information costs described above. The optimal tax rate is acyclical* ($\varphi^* = 1$) *if* $\eta = 1$, *procyclical* ($\varphi^* < 1$) *if* $\eta < 1$, *and countercyclical* ($\varphi^* > 1$) *if* $\eta > 1$.

The intuition is as follows. Because (and only because) of the need to raise tax revenue, it can be optimal to distort the agents' attention choices away from their laissez-faire counterparts. By an envelope argument, the first-order effect of such a distortion on consumer surplus is zero. To understand

[17] The assumption that the payoff is given by the product rather than the sum of expected utility and attention costs does not change the essence; it only facilitates a simple optimality condition for ρ, in the form of condition (**??**) in the Appendix. The scalar $-(1-\frac{1}{\eta})$, on the other hand, accounts for the fact that $u(x)$ and thus also the optimal $\mathbb{E}\left[u(x) - qx\right]$ flips sign as η crosses 1; rescaling $K(\rho)$ by that scalar thus makes sure that attention is indeed costly, regardless of η.

the planner's choice of φ, we can therefore focus on its effect on tax revenue. Using (25) to compute the resulting tax revenue, we can verify the following property: starting from the value of ρ that obtains when the tax is state-independent, an increase in ρ raises tax revenue when $\eta > 1$ and lowers it when $\eta < 1$. It follows that the planner wishes to stabilize aftertax prices so as to stimulate attention when $\eta > 1$, and to do the exact opposite when $\eta < 1$. In both cases, the planner therefore aims at higher tax revenue; but the tax cyclicality that best serves this goal flips sign as η crosses 1.

To conclude, this example has illustrated two insights from Theorem 2 and Corollary 2: first, how optimal attention naturally depends on the variance of, or uncertainty in, prices (Proposition 3); and second, how this in turn justifies state-dependency in taxes as an instrument for regulating attention (Proposition 4). At the same time, this example has shown that the optimal sign of such state-dependency is ambiguous in general. Furthermore, if practical or other considerations outside the model preclude the planner from using state-dependent taxes, this margin becomes moot and we are effectively back to the strongest version of our irrelevance result: once φ is restricted to 1, the planner lacks the ability to regulate ρ, and the optimal tax reduces to $\frac{\tau}{q} = \frac{\Lambda}{\eta\rho}$, echoing our discussion of equation (25). This further substantiates our point that, although the optimality of state-dependent taxes is a novel and valuable *theoretical* insight, our irrelevance result remains the main *practical* takeaway.

3.6 Adding production

We now extend our analysis to environments with endogenous production and endogenous government spending. We briefly introduce the environment, verify that our irrelevance results go through, and discuss why this is true, deferring the formal details to Appendix **??**.

Consider first production. This is represented by a technological constraint $H(\boldsymbol{x}^S, y^S, \theta) \leq 0$ for each θ, where $(\boldsymbol{x}^S, y^S)$ denotes the production vector and H is increasing and concave in it. This technology is operated by a representative, fully-attentive, competitive firm. For each θ, this firm thus maximizes profits, $\boldsymbol{p} \cdot \boldsymbol{x}^S + y^S$, subject to $H(\boldsymbol{x}^S, y^S, \theta) \leq 0$.[18] Consider next government spending. This takes the form of purchasing a vector $\boldsymbol{g} \in \mathbb{R}_+^N$ of taxable goods from that firm. These goods in turn convey public services: the utility of type$-h$ agents in state θ is now $u^h(\boldsymbol{x}^h, y^h, \theta) + \tilde{u}^h(\boldsymbol{g}, \theta)$, for some $\tilde{u}^h$ that is increasing and concave in $\boldsymbol{g}$. Finally, total government spending, $G = \boldsymbol{p} \cdot \boldsymbol{g}$, must be financed, in each state, by the same taxes as in our previous analysis.[19]

[18] Note that production is separable across state. This simplifies the exposition, without driving Proposition 5 below.

[19] Note that the government does not directly engage in production. This is without any loss of generality: government production can be replicated by market production by having the government hire a competitive firm to run its technology.

In this economy, the Ramsey planner regulates both consumer and producer prices. Invariance, though, guarantees that consumers can "see through" both kinds of prices. Along with the fact that the planner internalizes the consumers' attention costs, this guarantees that our earlier results about optimal taxes go through.

Proposition 5. *Consider the extension with endogenous production and government spending described above. Theorems 1 and 2 continue to hold, and so do Corollaries 1 and 2.*

Compared to our main analysis, the only difference is that the values of $\boldsymbol{p}_\theta$ and λ_θ that enter the optimal tax formulas are now endogenous to the Ramsey allocation. Intuitively, prices are determined by both demand and supply, while the shadow value of taxes reflects the social value of public goods. Formally, the tax formulas remain unchanged, and are simply complemented with the following additional conditions, which together characterize the entire solution to the planner's problem:[20]

$$\boldsymbol{p}_\theta = \frac{\nabla_{\boldsymbol{x}} H(\boldsymbol{x}_\theta^S, y_\theta^S)}{\nabla_y H(\boldsymbol{x}_\theta^S, y_\theta^S)}, \quad \lambda_\theta \boldsymbol{p}_\theta = \sum_{h\in\mathcal{H}} \mu^h \beta^h \nabla_{\boldsymbol{g}} \tilde{u}^h(\boldsymbol{g}_\theta, \theta), \quad (\boldsymbol{q}_\theta - \boldsymbol{p}_\theta)\cdot \sum_{h\in\mathcal{H}} \mu^h \boldsymbol{x}_\theta^h = \boldsymbol{p}_\theta \cdot \boldsymbol{g}_\theta, \quad \begin{pmatrix} \boldsymbol{x}_\theta^S \\ y_\theta^S \end{pmatrix} = \sum_{h\in\mathcal{H}} \mu^h \begin{pmatrix} \boldsymbol{x}_\theta^h \\ y_\theta^h \end{pmatrix} + \begin{pmatrix} \boldsymbol{g}_\theta \\ 0 \end{pmatrix}.$$

The first equation gives production efficiency (also, firm optimality); the second gives the optimal provision of public goods; the third gives the government budget; the last one gives market clearing.

While Proposition 5 is intuitive and straightforward, it is important to note that it relies on invariance, not only because Theorems 1 and 2 themselves invoke invariance, but also because of the following, subtler, reason. By modeling the production sector with a single representative firm, we have effectively presumed production efficiency: there is no social gain from the use of intermediate-input taxes or firm-specific prices. Under the two key assumptions about attention made here—full attention on the firm side plus invariant attention costs on the consumer side—this property follows in our context by the exact same argument as in Diamond and Mirrlees (1971a). But if we relax *either* of these assumptions, production efficiency may fail. We return to this point in Section 5.

Finally, the assumption that the utility of the public goods $\boldsymbol{g}$ is additively separable from the utility of the private goods is also without any loss of generality: it only simplifies the exposition, making sure that $\boldsymbol{g}$ does not become an input in the demand functions.

[20] We assume an interior solution to the planner's problem. Otherwise, the optimality conditions may have to be replaced with the usual Kuhn-Tucker conditions.

4 Relaxing invariance

We now discuss how our irrelevance result breaks when attention costs fail to be invariant. The basic idea is that violations of invariance open the door to cognitive externalities a la Angeletos and Sastry (2025): attention costs depend on the volatility or other statistical properties of aftertax prices (e.g., their "complexity" or "salience"), which consumers take as given but the government can regulate through its tax policy. This in turn calls for taxes to serve not only the Ramsey function of the previous sections, but also a novel, Pigou-like, corrective function: as tools to simplify the price system and reduce the agent's cognitive burden. In this section, we characterize this new corrective function and go on to show how it translates to a complementary rationale of state-dependent taxes.

4.1 A new corrective function for taxes: easing attention costs

To see what this new corrective role of taxes is and how it enters our optimal tax formulas, we return to the main setting of Section 3 (i.e., without production) and revisit the indirect utility and the demand of the fictitious attentive economy developed there. When attention costs were invariant, the dependence of these objects on ϕ^* vanished. Now, this is no more true: ϕ^* acts like an endogenous taste parameter, which not only influences consumer surplus and tax revenue, but also depends on the distribution of prices system and thus on the tax system. Consider first how this matters for tax revenue. To account for the indirect effect of $\boldsymbol{Q}$ via ϕ^*, we simply redefine the Slutsky matrices as

$$\boldsymbol{S}^h_{\theta,\theta'} = \nabla_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_{\theta'} + \left(\partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'}\right)\left(\bar{\boldsymbol{x}}^h_\theta\right)^\top, \tag{26}$$

where $\nabla_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \left(\frac{dX^h_i\left(\theta', \boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q})\right)}{dq_{\theta,j}}\right)_{i,j\in\mathscr{I}}$ is now the matrix of *total* price derivatives, inclusive of the effect via ϕ^*, and $\partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \frac{\partial \boldsymbol{X}^h\left(\theta', \boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q})\right)}{\partial w_\theta}$ is the vector of income effects. Although these Slutsky matrices are different from those featured in Theorems 1 and 2, the common thread is that prices affect demand not only through the classical substitution effects, but also through the response of attention to prices—and what matters for tax revenue is just the overall compensated demand elasticities, not the precise mechanism "under the hood." A more material difference emerges when we consider consumer surplus. This is now the sum of two terms:

$$\frac{d}{d\boldsymbol{q_\theta}} V^h\left(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*(\boldsymbol{Q})\right) = -\underbrace{\partial_{w_\theta} V^h \bar{\boldsymbol{x}}^h_\theta}_{\text{Standard effect}} - \underbrace{\partial_{\boldsymbol{q}_\theta} C^h\left(F^h, \phi^*(\boldsymbol{Q})\right)}_{\text{Cognitive externality}}\Big|_{F^h = F^h(\phi^*(\boldsymbol{Q}))}.$$

The first term is the familiar one: it captures the welfare loss from the distortion of consumption, re-expressed as a multiple of $\bar{\boldsymbol{x}}^h_\theta$ using Roy's identity, and accounting—again "under the hood"—the response of optimal attention. The second term is new: it captures the dependence of attention costs on the complexity or the stochastic properties of the price system. We therefore reach the following result, which extends Theorem 2 away from invariant attention costs.

Theorem 3. *For general (possibly non-invariant) attention costs, the optimal taxes satisfy:*

$$\sum_{h\in\mathcal{H}} \mu^h \left[(\lambda_\theta - \gamma^h_\theta)\bar{\boldsymbol{x}}^h_\theta + \sum_{\theta'\in\Theta} \lambda_{\theta'} \left(\boldsymbol{S}^h_{\theta,\theta'}\right)^\top \boldsymbol{\tau}_{\theta'} + \beta^h\, \Xi^h_\theta \right] = \boldsymbol{0} \qquad \forall\theta, \tag{27}$$

where $\Xi^h_\theta \equiv -\frac{\partial}{\partial \boldsymbol{q}_\theta} C^h\big(F^h, \phi^*(\boldsymbol{Q})\big)\Big|_{F^h = F^h(\phi^*(\boldsymbol{Q}))}$ *is the cognitive externality experienced by type* h *in state* θ.

Relative to Theorems 1 and 2, the optimal taxes now contain a Pigouvian correction, equal to $\sum_{h\in\mathcal{H}} \mu^h \beta^h \Xi^h_\theta$, the Pareto-weighted average cognitive externality.[21] To make this new term more tangible, abstract once again from distributional concerns and from uncertainty other than in producer prices, as we did before in Corollaries 1 and 2, but now let attention costs be neither state-separable nor invariant. Suppose further that the effect of the price system on attention costs is summarized in the same set of moments as that entering demand and tax revenue—and, with the same abuse of notation as in those corollaries, rewrite demand as $\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma})$ and attention costs as $C^h(F,\phi) = C^h(F,\boldsymbol{\sigma})$. We can then recast Theorem 3 as follows.

Corollary 3. *Consider the class of economies described above, where all uncertainty is about prices and its effect on demand and attention costs in summarized in* $\boldsymbol{\sigma}$*. The optimal taxes then satisfy*

$$\boldsymbol{S}(\boldsymbol{q})^\top \boldsymbol{\tau} + \textstyle\sum_k \mathbb{E}[\partial_{\sigma_k} \hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma})\boldsymbol{\tau}]\partial_{\boldsymbol{q}} m'_k(\boldsymbol{q}) = -\Lambda \hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma}) + \sum_k \mathbb{E}\big[\sum_h \mu^h \beta^h \partial_{\sigma_k} C^h(F,\sigma)\boldsymbol{\tau}\big]\partial_{\boldsymbol{q}} m'_k(\boldsymbol{q}). \tag{28}$$

The two terms in the left hand side of this tax formula capture the effect of changing $\boldsymbol{\tau}$ on tax revenue. The first term on the right hand side capture the standard effect on consumer surplus. The third term capture the cognitive externality. Altogether, the optimal taxes therefore combined the Ramsey function of our preceding analysis with the new, Pigou-like function implied by the violation of invariance. We next illustrate this function in an example where agents make relatively larger mistakes

[21] In an economy with endogenous production (Section 3.6), violations of consumer invariance create an additional channel. If producer prices enter the consumers' cognitive state, so that $z = (\theta, \boldsymbol{p}, \boldsymbol{q})$, then changes in the aggregate production plan shift consumers' priors and alter expected welfare and demand. The planner may therefore target a different point on the production frontier in order to influence producer prices, even though the firm itself continues to produce efficiently. See Appendix **??** for details.

when they live in a more volatile, or more complex, market environment. We show, in particular, that this function translates to an additional rationale for state-dependance in the optimal tax. Violations of invariance thus lead to the same policy recommendation as violations of state-separability, despite the different underlying motivations (cognitive externalities here vs tax revenue collection before).

4.2 An example with variance-reduction costs

We revisit the example of Section 3.5, making two changes. First, we impose $\eta = 1$ to shut down the effects of non-separability (already characterized in Proposition 4) and isolate instead the effects of non-invariance. Second, we introduce non-invariance by letting the costs of attention depend, not only on the mutual information between prices and signals, but also on the volatility of prices.

To motivate this change, consider momentarily an economy where attention costs are given by reduction in the agents' posterior uncertainty about prices, with uncertainty measured by variance:

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

for some increasing and convex $\tilde{K}$. Preserving the notation of Section 3.5, let $v = \mathbb{V}(\log q)$ and let ρ be the square correlation between ω and q. We then have $\mathbb{V}[\log q \mid \omega] = (1-\rho)\mathbb{V}[\log q]$ and can thus rewrite the attention cost as $\tilde{K}(\rho v)$. Seen from this prism, the key novelty relative to Section 3.5 is the emergence of v as an additional input in attention costs—this is how non-invariance matters, and this is where the cognitive externality comes into play.

Let us now generalize the above example by letting

$$\text{attention cost} = K(\rho, v),$$

for some function K that depends flexibly on ρ and v. Mutual-information costs map to $K_v = 0$ and preclude cognitive externality. A cognitive externality instead emerges whenever $K_v \neq 0$. The case of $K_v > 0$ nests the above example and captures the idea that the cognitive burden of tracking prices and choosing the right consumption is higher when prices are more volatile. The opposite case of $K_v < 0$ could capture a situation in which consumers have *easier* time discerning, and optimally responding to, price changes when these changes are larger. To simplify the exposition, we finally assume that K is such that the optimal ρ is interior and satisfies $\partial \log \rho / \partial v > -1$.

The planner can regulated v by choosing $\varphi \neq 1$. In Section 3.5, the only rationale for doing so was to raise more tax revenue. That rationale has been switched off here by restricting $\eta = 1$, but a new rationale has obtained because of the violation of invariance and the resulting cognitive externality. The next result then shows how the planner leverages the cyclicality of the tax to ease that externality.

Proposition 6. *Consider the economy described above. The optimal tax rate is procyclical if* $K_v > 0$ *(as in the example with variance-reduction costs) and it is countercyclical if* $K_v < 0$*. In particular:*

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

where $\rho(v)$ *is the optimal attention as a function of the aftertax price volatility,* $\epsilon(v)$ *is the corresponding elasticity, and* $K_v(\rho, v)$ *is the relevant cognitive externality.*

Proposition 6 shows that non-invariant costs, just like non-separable costs, can introduce state-dependence in the optimal tax. But as already explained, the rationale is different. In Proposition 4, state-dependent taxes helped raise more tax revenue; here, they help ease consumer attention. When $K_v > 0$, this goal is achieved by reducing the volatility in aftertax prices, which in turn is achieved with countercyclical taxes. When $K_v < 0$, the same goal is achieved with procyclical taxes. Proposition 6 further shows that the absolute *magnitude* of the optimal cyclicality is inversely related to the elasticity of attention. This reflects the optimal balance between the Ramsey and Pigou goals. Finally, note that the mapping from φ to the volatility in consumer prices is "mechanical" here because supply is infinitely elastic (pretax producer prices are exogenously specified), but the insights generalize as long as supply is not completely inelastic. In this case, the endogenous adjustment in p can offset the policy effect on q, but this offsetting is only partial, so the essence remains unchanged.

We close this section by illustrating the combined insights of Propositions 4 and 6. In particular, we let $K(\rho, v) = \frac{\kappa}{2}\left(\mathbb{V}[\log q] - \mathbb{E}\left[\mathbb{V}[\log q \mid \omega]\right]\right)^2 = \frac{\kappa}{2}(\rho v)^2$ for $\kappa > 0$, as in the example from the beginning of this section;[22] but now we also allow $\eta \neq 1$, and we go on to draw, in Figure 1, the optimal φ as a function of κ and of η. The solid blue line corresponds to $\eta = 1$ and isolates the effect of non-invariance: echoing Proposition 6, the optimal tax system in this case is procyclical ($\varphi^* < 1$) because and only because $K_v > 0$. The two other curves let $\eta \neq 1$, thus interacting the effects of non-invariance and non-separability. 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.

[22] This example is not directly nested in Proposition 6, because 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.

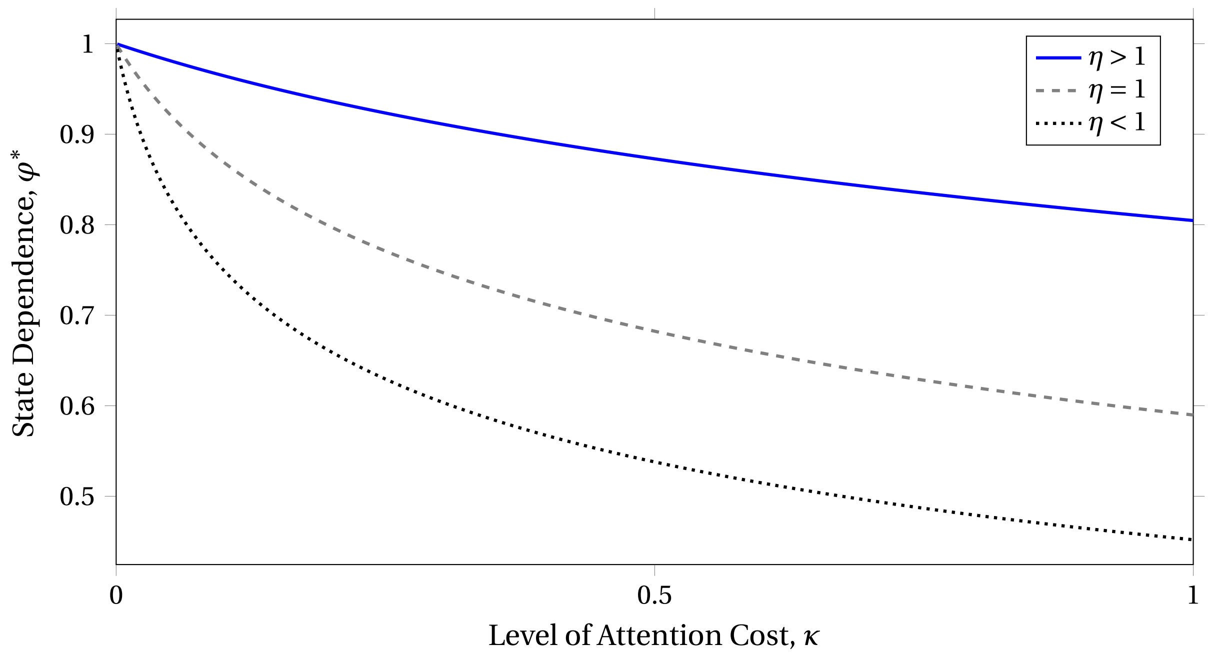

Figure 1: Optimal state dependence φ^* with non-invariant costs.

5 Discussion

We conclude by discussing the precise meaning of our results, the differences between "rational" and "behavioral" inattention, and a few practical takeaways.

Key takeaways: irrelevance and state-dependence

The main insight of our analysis is an irrelevance result: when agents are rationally inattentive, a benevolent planner need not correct the effects of inattention on allocation; need not try to separate inattention from preferences; and may thus design optimal taxes using the same sufficient-statistic tax formulas as those developed in classic public finance. The strongest version of this result (Theorem 1 and Corollary 1) was obtained by under two conditions on attention costs: invariance and state-separability. From a conceptual perspective, violations of these assumptions played a very different role: the one accommodated rational confusion without upsetting the logic of our irrelevance result; the other upset that logic and introduced a novel reason for corrective taxation. Violations of these assumptions nevertheless ended up being close cousins in the following sense: they both translated to a rationale for state-dependence in the optimal taxes. In the one case, state-dependence helped attain a better balance between collecting tax revenue and distorting consumption; in the other, it helped ease people's cognitive burden. This two-part rationale for state-dependent taxes thus emerged as the second main insight of our analysis.

Volatility, complexity, and sparsity.

In Section 4.2, we informally equated the volatility of consumer prices with the "complexity" of the price system. This interpretation can be expanded by defining the complexity of the price system as its entropy—and by letting this entropy increase attention costs holding constant the mutual information of prices and signals. Formally, let attention costs be $K(\mathscr{I}, \mathscr{E})$, where $\mathscr{I}$ is the mutual information between the price vector $\boldsymbol{q}$ and the signal ω and $\mathscr{E}$ is the entropy of $\boldsymbol{q}$. This nests the setting of Section 4.2, with $\mathscr{I}$ and $\mathscr{E}$ in place of ρ and v, respectively, and allows one to operationalize that section's ideas beyond the log-normal, one-good example.

The following variant example then helps further illustrate what we have in mind—and also connect to our earlier discussion about sparsity. Suppose that there are 10 exogenous states of nature, corresponding to 10 different realizations of the underlying producer prices. Suppose next that consumers prefer "sparsity" or "simplicity" in a dual sense: first, their attention costs decrease by pooling their demands across different cognitive states (i.e., by choosing the same F conditional on multiple z); and second, their attention cost decrease when they have to contemplate fewer "cognitive states" in the first place (i.e., when the size of the support of z is smaller). The first notion opens the door to "sparse" behavior and "coarse" demands, as in Gabaix (2014) and Gul et al. (2017): the consumer may optimally pool demand across multiple price realizations. The second notion calls for the planner to simplify or coarsen the price system itself: the planner may optimally use state-dependent taxes to map the 10 producer-price realizations to, say, only 3 consumer-price realizations ("normal", "abnormally low", and "abnormally high").

Connections to behavioral public finance

The last example illustrates, not only one can accommodate "sparsity" under the rational-inattention umbrella, but also how this can lead to starkly different *normative* conclusions than the behavioral alternative put forward in Gabaix (2025) and Farhi and Gabaix (2020). In both our framework and theirs, sparsity and inattention are the product of cognitive frictions, modeled as the cost of choosing a perception close to the truth. But whereas the planner in our paper fully internalizes these costs when assessing the efficiency of market outcomes and when designing the optimal taxes, the planner in those papers disregards these costs entirely—put simply, it is *as if* the planner disapproves of the agents' preferences. It is this paternalistic view of inattention that drives the corrective taxes advocated for in those papers.

To illustrate this point further, revisit the aforementioned example and drop the second of the two senses in consumers preferred "sparsity" or "simplicity": consumers may still economize attention costs by pooling their demands across multiple price realizations, but they no more gain from a coarsening of the price system. In this case, agents still exhibit sparse behavior; but attention costs are invariant, so Theorem 2 applies: optimal taxes do *not* contain a corrective, Pigou-like term. And while a violation of invariance *does* introduce a corrective goal for taxes, this goal is very different than that found in Farhi and Gabaix (2020): the planner aims, not at undoing the effects of inattention, but rather at coarsening the price system and easing people's cognitive burden.

Cognition, invariance, and salience When we specified how the consumer reasoned and behaved, we introduce the notion of the cognitive state $z = (q, \theta)$ and distinguished from the underlying physical state θ. We then explained that, although q would ultimately be a transformation of θ in equilibrium, the distinction was meaningful in general because inattentive agents may have difficulty seeing through different transformations of the same underlying uncertainty—and we then went on to identified invariance as the condition on cognition that made sure that this distinction was immaterial.

What if we include additional, seemingly-superfluous, variables in the cognitive state? In particular, what if we let z include taxes $\boldsymbol{\tau}$ and producer prices $\boldsymbol{p}$ on top of consumer prices $\boldsymbol{q}$? Fully attentive agents do not care about $\boldsymbol{p}$ and $\boldsymbol{\tau}$ separately from $\boldsymbol{q}$, and they perfectly know $\boldsymbol{q}$. When agents are inattentive, they may not know $\boldsymbol{q}$ perfectly. Nonetheless, because $\boldsymbol{q}$, $\boldsymbol{p}$, and $\boldsymbol{\tau}$ are all deterministic transformations of θ, and because agents can costlessly "see through" such transformation when attention costs are invariant, the following is true: under invariant costs, there is no change in our results as we expand the cognitive state from $z = (\theta, \boldsymbol{q})$ to $z = (\theta, \boldsymbol{q}, \boldsymbol{p}, \boldsymbol{\tau})$. If instead invariance fails, this distinction can not only matter but also help capture new ideas.

To illustrate, consider the role of "salience". The following notion of salience can be captured even without the above revision for z : if large price changes are more salient than small price changes in the sense that they are easier to detect, then the planner may wish to induce larger changes in prices in order to ease people's attention costs. This maps, basically, to $K_v < 0$ and $\varphi^* > 1$ in the context of Section 4. But there is a different notion of salience, which can be captured once we have revised z along the above lines. In this case, taxes could be less salient than prices in the following sense: the costs of learning about $\boldsymbol{\tau}$ are larger than the costs of learning about $\boldsymbol{p}$, and the consumer cannot just sidestep the problem by learning directly about $\boldsymbol{q}$ at the lowest cost. This amounts to a particular violation of invariance.

Revisit now the evidence that consumers are relatively irresponsive to non-salient changes in taxes (Chetty et al., 2009). Under our paper's lens, this evidence points towards the aforementioned violation of invariance. It also suggests that an unexpected change in τ may have less distortionary effects than a change in the mean or "steady state" value of τ. This in turn can justify a particular state-dependence in taxes—taxes increasing relatively steeply when there is an unexpected need for tax revenue—but may not necessarily justify higher tax on average. We leave a further exploration of these insights for future work.

6 Conclusion

This paper studied optimal commodity taxation when consumers are rationally inattentive. Our main finding was an irrelevance result: under certain conditions (invariant and state-separable attention costs), policy makers should neither try to undo the effects of inattention nor reconsider the lessons of classical public finance; instead, the can design taxes *as if* agents were fully attentive. We then showed that violations of those conditions provided two complementary rationales for making taxes state-dependent, without however justifying a systematic shift towards either lower or higher taxes.

These findings qualify the normative arguments found in an emerging behavioral public finance literature. Behaviors akin to sparsity, narrow bracketing, and mental accounting are all consistent with invariant attention costs; they therefore do not call for corrective measures by themselves. And while a corrective rationale emerges when attention costs violate invariance, this rationale takes a much subtler form than that found in that literature: the planner should not just "undo" the effect of inattention on allocations; rather, the planner should use state-dependent taxes so as to ease people's cognitive burden.

We illustrated this new rationale for corrective state-dependent taxes in an example. The basic idea is that the planner should use such state-dependent taxes to stabilize prices or coarsen the price system when agents get confused by volatile prices or complex markets, while it should do the opposite if agents have easier time discerning larger price changes. And while this insight is too speculative at this point to be of practical value, it deserves further attention.

Finally, our paper followed the Ramsey tradition on commodity taxation, presuming that some goods cannot be taxed and restricting the taxes on all other goods to be linear (i.e., proportional to the amount consumed or supplied). We thus left outside the purview of our analysis the question of how rational inattention affects the optimal non-linear taxation of labor income, as in the Mirrlees tradition. In that context, our present insight about state-dependent commodity taxes may translate to a new rationale for non-linear income taxation. Furthermore, to the extend that the noise due to inattention does not wash out across multiple choices, optimal income taxes may also aim at providing insurance against that noise. We leave the exploration of these ideas to future work.

A Appendix

To be added

References

Angeletos, George-Marios and Alessandro Pavan, "Policy with Dispersed Information," *Journal of the European Economic Association*, 2009, *7* (1), 11–60. 1

— **and Karthik A Sastry**, "Inattentive economies," *Journal of Political Economy*, 2025, *133* (7), 000–000. 1, 1, 3, 8, 11, 13, 4

Boccanfuso, Jérémy and Antoine Ferey, "Inattention and the Taxation Bias," *Journal of the European Economic Association*, October 2023, p. jvad056. 1

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, July 2015, *105* (7), 2183–2203. 1, 1, 2.1, 2.1

— , — **, and John Leahy**, "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, May 2019, *86* (3), 1061–1094. 1

— , — **, and** — , "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," *Journal of Political Economy*, June 2022, *130* (6), 1676–1715. Publisher: The University of Chicago Press. 1, 2.1, 3, 13

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, August 2009, *99* (4), 1145–1177. 1, 5

Correia, Isabel, Juan Pablo Nicolini, and Pedro Teles, "Optimal Fiscal and Monetary Policy: Equivalence Results," *Journal of Political Economy*, 2008, *116* (1), 141–170. 1

Diamond, P. A., "A many-person Ramsey tax rule," *Journal of Public Economics*, November 1975, *4* (4), 335–342. 1, 1, 3, 3.3, 3.3

Diamond, PA and A Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 1971, *61*, 8–27. 1, 3, 3.6

— **and** — , "Optimal taxation and public production II: Tax rules," *American Economic Review*, 1971, *61*, 261–278. 1, 1, 3, 3.3

Farhi, Emmanuel and Xavier Gabaix, "Optimal taxation with behavioral agents," *American Economic Review*, 2020, *110* (1), 298–336. 1, 1, 2.1, 2.5, 5

Flynn, Joel P and Karthik A Sastry, "Strategic mistakes," *Journal of Economic Theory*, 2023, *212*, 105704. 1, 1, 2.1, 2.1, 3, 3.2, 3.2

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki, "Stochastic choice and revealed perturbed utility," *Econometrica*, 2015, *83* (6), 2371–2409. 2.1

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710. 2.1, 5

_ , "A Theory of Complexity Aversion," *Available at SSRN*, 2025. 5

Galí, Jordi, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press, 2008. 1

Gul, Faruk, Wolfgang Pesendorfer, and Tomasz Strzalecki, "Coarse Competitive Equilibrium and Extreme Prices," *American Economic Review*, January 2017, *107* (1), 109–37. 5

Hébert, Benjamin and Jennifer La'O, "Information acquisition, efficiency, and nonfundamental volatility," *Journal of Political Economy*, 2023, *131* (10), 2666–2723. 1, 3, 11, 13

Kőszegi, Botond and Filip Matějka, "Choice simplification: A theory of mental budgeting and naive diversification," *The Quarterly Journal of Economics*, 2020, *135* (2), 1153–1207. 1, 1, 3.3, 3.3, 3.4

Lian, Chen, "A Theory of Narrow Thinking," *The Review of Economic Studies*, 2021, *88* (5), 2344–2374. 1, 1, 3.2, 3.3, 3.3, 3.4

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, March 2023, *61* (1), 226–273. 1

Matějka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–298. 1, 1, 2.1, 2.1

Matějka, Filip and Christopher A Sims, "Discrete Actions in Information-Constrained Tracking Problems," *CERGE-EI mimeo*, 2011. 1, 1

Ramsey, F. P., "A Contribution to the Theory of Taxation," *The Economic Journal*, 1927, *37* (145), 47–61. 1, 2.1, 2.4

Sims, Christopher A., "Implications of rational inattention," *Journal of Monetary Economics*, April 2003, *50* (3), 665–690. 1, 1, 2.1, 2.1

Stahl, Dale O, "Entropy control costs and entropic equilibria," *International Journal of Game Theory*, 1990, *19*, 129–138. 2.1, 2.1

Stevens, Luminita, "Coarse Pricing Policies," *Federal Reserve Bank of Minneapolis mimeo*, 2015. 1

Taubinsky, Dmitry and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *The Review of Economic Studies*, 11 2017, *85* (4), 2462–2496. 1

Woodford, Michael, "Modeling imprecision in perception, valuation, and choice," *Annual Review of Economics*, 2020, *12*, 579–601. 2.1