Diamond-Mirrlees meets Sims: Optimal Taxation with Rational Inattention

George-Marios Angeletos Matias Bayas-Erazo

July 10, 2026

Abstract

We study optimal commodity taxation when consumers are rationally inattentive. In our framework, consumers may be inattentive to prices and taxes, may allocate attention unevenly across decisions, and may display behavior resembling sparsity or mental accounting. Our main result is that this need not change optimal tax design. In a useful benchmark, a benevolent planner sets taxes *exactly* as in classical public finance: the key sufficient statistics are the reduced-form demand elasticities, regardless of the cognitive mechanisms behind them. Away from this benchmark, our analysis offers two complementary rationales for state-dependent taxes—easing the tax distortion and/or simplifying the agents' attention problem—but does not justify either a blunt correction for inattention or a special consideration of tax salience. The substantive assumption that drives these lessons, and that explains their difference from a recent literature on behavioral public finance, is that inattention is modeled as a rational choice.

1 Introduction

How should commodity taxes be designed when consumers are inattentive? One 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 (e.g., Chetty et al. 2009; Goldin 2015; Allcott et al. 2018; Farhi and Gabaix 2020). This paper takes a different route. We ask how optimal commodity taxation changes 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 face uncertainty about these objects and choose stochastic consumption rules, 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 finding is an *irrelevance* result: taxes should be designed as if agents were fully attentive. The sharpest version of this result is obtained under two conditions on attention costs—invariance and state-separability—whose meaning and role we discuss below. In this benchmark, the classic sufficient-statistics logic of Ramsey (1927), Diamond and Mirrlees (1971b) and Diamond (1975) survives intact: for tax design, what matters are the compensated price elasticities of the reduced-form demands, not the underlying preferences or cognitive mechanisms. Put simply, a benevolent planner should not care what lies "under the hood" of observed demand.

This irrelevance result is, in our view, a useful starting point. A real-world tax authority does not observe people's tastes, 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 after-tax 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 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.

As already noted, this result makes two assumptions on attention costs. The first 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. The second 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. Together, these assumptions imply that the Ramsey problem in our inattentive economy translates, state by state, to a classical Ramsey problem with 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 neither ad hoc nor empty. 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, even under these restrictive assumptions, rational inattention can produce systematic underreaction in some markets, stronger responses in others, and patterns that resemble coarse thinking or mental accounting. The key insight that emerges from our benchmark is that, once cast as the product of rational inattention, such behaviors do not upset classical taxation results. A complementary insight is that, contrary to conventional wisdom, the positive effects of inattention are ambiguous: when inattention is rational and thus responsive to incentives, the reduced-form demand may be *more* elastic than in the frictionless counterpart, justifying lower taxes.[1]

State-dependent taxes. Away from our benchmark, the following high-level insights survive: rational inattention is akin to appropriately modified preferences; and the Ramsey tradeoff between raising more tax revenue and reducing consumer surplus is still governed by certain reduced-form elasticities. What changes, though, is the precise definition and measurement of these elasticities, and the scope for a Pigouvian correction. The upshot is a dual rationale for state-dependent taxes.

First, suppose invariance holds but state separability fails, the case pertinent to mutual-information

[1] That said, our irrelevance result says that the planner does not even need to know what these positive effects are: to design taxes, the planner should just look at the reduced-form elasticities.

costs. This continues to rule out a Pigouvian correction: there are no externalities, or wedges, due to inattention. It also preserves the equivalence of our Ramsey problem to that of a counterfactual, fully-attentive, economy with appropriately modified preferences. But these as-if preferences and the resulting demand systems are no more separable across states, even if the primitive preferences were separable to start with. Intuitively, because consumers confuse different realizations of uncertainty, the level of demand at any given aftertax price depends on what the taxes are in all other price points. Furthermore, because the optimal level of attention depends on the uncertainty consumers face in aftertax prices, and because the planner can regulate this uncertainty by letting taxes vary across state realizations, the planner can in principle use this lever to improve the Ramsey tradeoff—extract more tax revenue with less sacrifice in consumer surplus.

Second, suppose invariance fails. In this case, attention costs depend, not only on the underlying primitive uncertainty (i.e., the exogenous uncertainty in preferences and technologies), but also on how this uncertainty is translated via the price system. Concretely, a failure of invariance can accommodate a scenario where consumers get more confused when prices are more volatile, or where they have a preference for simplicity in the sense of a coarser price system. Angeletos and Sastry (2025) have shown that this amounts to a "cognitive externality" and a failure of the first welfare theorem. Here, we show that the associated Pigouvian correction translates to an additional rationale for state-dependent taxes: by regulating the uncertainty in prices, the planner cannot only ease her Ramsey tradeoff but also simplify the agents' attention problem.

Tax salience. With mutual-information or, more generally, invariant attention costs, decomposing a consumer price into a producer price and a tax is only a relabeling, which leaves demand and consumer surplus entirely unchanged. Away from that benchmark, consumers may instead be differentially attentive to pretax and aftertax prices. A failure of invariance thus allows us to revisit the results of Chetty et al. (2009) and others on "tax salience" under the rational-inattention prism.

We start with the special case in which producer prices are exogenous to tax policy, because supply is perfectly elastic. In this case, saying that taxes are non-salient is equivalent to saying that consumers are inattentive to aftertax prices, which is already accommodated in the preceding results. It follows that tax salience can play a distinct role under rational inattention only insofar as producer prices adjust in equilibrium—an insight that by itself is new to the literature.

Turning to that more general scenario, with imperfectly elastic supply, we observe that taxes then affect demand via *both* pretax and aftertax prices. Intuitively, if higher taxes depress demand and

thus producer prices, and if consumers are more attentive to pretax prices than to aftertax prices, then demand may fall less than in a counterfactual scenario with salient taxes, easing the Ramsey tradeoff. From this perspective, optimal taxes depend on the degree of tax salience along with the elasticities of demand and supply, which appears to invalidate the classical tax formulas. Nonetheless, we show that those formulas continue to hold if the relevant elasticity is identified by an economy-wide tax reform as opposed to an agent-specific price hike: the *equilibrium* response of demand to a tax reform accurately summarizes all the underlying mechanisms relevant for the Ramsey tradeoff, and the planner can once again avoid looking "under the hood".

Discussion. To recap, our paper has two main lessons. The first—irrelevance—is that inattention to taxes, sparsity, mental accounting, and related "non-classical" behaviors do *not* necessarily invalidate classical taxation results. The second is a novel, dual rationale for state-dependent taxes, along with a rejection of a blunt Pigouvian correction for "behavioral wedges".

The substantive assumption that drives these two lessons, and that explains their contrast to those found in a recent literature on behavioral public finance, is that inattention is modeled as the product of rational choice subject to attention costs, which a benevolent planner ought to internalize. The two lessons then 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. Put differently, we first explain why, contrary to the prevailing view, the available evidence on inattention and tax salience does not necessarily have a bite on optimal tax design. We then bring to light a novel mechanism, whose quantitative importance remains unexplored. Absent any evidence speaking specifically to this mechanism, or to a true *ir*rationality in consumer behavior, the main practical takeaway from analysis is the irrelevance result.

Related literature. The rational-inattention framework pioneered by Sims (2003) has found a wide range of applications. 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 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 its 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 ease the agents' cognitive burden.[2]

Last but not least, our paper connects to a growing literature on behavioral public finance. Chetty et al. (2009) provide the foundational evidence that consumers underreact to non-salient taxes. Motivated by this evidence, a large follow-up literature studies possible tax implications *outside* the rational-inattention framework (Goldin, 2015; Allcott et al., 2018; Mullainathan et al., 2012; Farhi and Gabaix, 2020; Taubinsky and Rees-Jones, 2017). Our non-paternalistic stance rules out the main corrective motive in much of this work. Instead, we identify conditions under which the usual demand-side statistics remain sufficient even when demand is generated by inattention. We also show that the evidence on tax salience can be understood, under the rational-inattention prism, as a violation of invariance and clarify how this changes, or not, optimal tax design. Finally, Boccanfuso and Ferey (2023) are close to us methodologically, but they study a different question: whether rational inattention can generate time inconsistency in optimal policy, and in particular a bias toward 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 discuss the main takeaways and the connection to behavioral public finance.

[2]Another example comes from Angeletos and Pavan (2009): in that paper, a state-contingent tax helps regulate the use of dispersed private information and thereby the aggregation of such information through endogenous public signals 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.

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 after-tax 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 may vary across states 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 together with tax policy determines 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 state θ, the producer price is p_θ, tax policy sets τ_θ, and the consumer price is $q_\theta = p_\theta + \tau_\theta$. In equilibrium, q will therefore 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 $F \in \mathcal{F}$ 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 to meet 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:

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

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 to 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:

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

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 on q, we can write the consumer problem as follows:

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

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 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 $\bar{x}(q,\phi) \equiv \int_{\mathscr{X}} x\, dF^*(x \mid q;\phi)$.[3] 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.[4] 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 after-tax price across the different

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

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

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.[5] 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 after-tax 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 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. The reason is that 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

[5] 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 in standard Ramsey models with public goods.

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 realized after-tax 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 demand is 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 problem as 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:

$$\hat{u}(\bar{x}) \equiv \max_{f\geq 0} \left\{ \int_{\mathcal{X}} \left(u(x) f(x) - K(f(x)) \right) dx \quad \text{s.t.} \quad \int_{\mathcal{X}} f(x)dx = 1, \quad \int_{\mathcal{X}} x f(x) dx = \bar{x} \right\} \quad \forall \bar{x} \tag{5}$$

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

As usual, concavity of $\hat{u}$ implies that the induced demand $\hat{x}(q) \equiv \arg\max_{\bar{x}} \left[\hat{u}(\bar{x}) - q\bar{x} \right]$ is decreasing in the price q. 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 price-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:

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

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. Appendix A.1 records the algebra.

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.[6] For each good i, the optimal tax then satisfies

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

where $\epsilon_i(q_i)$, 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.

[6] In the notation of the general model below, this is the one-type, quasilinear, good-by-good separable case, with $u(\boldsymbol{x}, y) = \sum_{i\in\mathscr{I}} u_i(x_i) + y$ and $C(F,\phi) = \sum_{\boldsymbol{q}\in\mathscr{Q}_\phi} \phi(\boldsymbol{q}) \sum_{i\in\mathscr{I}} \int_{\mathscr{X}} K_i(f_i(x_i \mid q_i))\, dx_i$, where $\boldsymbol{x} = (x_i)_{i\in\mathscr{I}}$ and $\mathscr{I}$ is the set of taxed goods.

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 isoelastic, with elasticity parameter $\eta > 0$. That is, $u(x) = \frac{x^{1-1/\eta}-1}{1-1/\eta}$ when $\eta \neq 1$, while $u(x) = \log x$ when $\eta = 1$. Realized consumption is $x = \bar{x}e^{\varepsilon}$, where $\bar{x}$ is the agent's target, $\varepsilon \sim \mathcal{N}(-\frac{1}{2}\nu^{-1}, \nu^{-1})$ is a random tremble, and ν is the precision with which this target is attained, or equivalently the level of attention.[7] The consumer chooses both $\bar{x}$ and ν conditional on q, so as 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 $\hat{u}(\bar{x}) \equiv \max_{\nu}\left[u(R(\nu)\bar{x}) - C(\nu)\right]$, 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 target $\bar{x}$ for given precision ν; the "outer" problem, on the left, trades off this benefit against the cost $C(\nu)$ to determine the optimal attention ν. Let $x^\circ = x^\circ(q, \nu)$ and $\nu^\circ = \nu^\circ(q)$ be the respective solutions of these problems. Then, the effective demand function is $\hat{x}(q) = x^\circ(q, \nu^\circ(q))$ and its elasticity is

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

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-adjusted return $R(\nu)$, while the choice of ν resembles a diversification strategy that reduces the riskiness of that portfolio at a cost $C(\nu)$. 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

[7] Note that $\mathbb{E}[x|\bar{x}] = \bar{x}\mathbb{E}[e^{\varepsilon}] = \bar{x}$ and $\mathbb{V}[\log x|\bar{x}] = \mathbb{V}[\varepsilon] = \nu^{-1}$, which justify these interpretations. Also note that the present example mirrors the additive costly-control example above, 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.

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 $\frac{\partial \log x^\circ}{\partial v} \frac{dv^\circ}{d \log q} < 0$, which proves that the elasticity is *higher* than its frictionless counterpart.[8] 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 4 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 allow the planner to completely sidestep the question of how inattention affects demand and, instead, design 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 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 conditions on attention costs—*state separability* à la Flynn and Sastry (2023) and *invariance* à la Angeletos and Sastry (2025), Caplin et al. (2022) and Hébert and La'O (2023). These

[8] Appendix A.2 gives the algebra behind this comparison and considers an extension in the spirit of Epstein-Zin showing that these signs are governed by the substitution parameter η, not risk aversion. At a corner, or more generally when the optimal attention choice does not respond locally to the price, the extra term in (11) can collapse. The conclusion is then weak: $\epsilon(q) \geq \eta$, and the optimal tax is weakly lower than its frictionless counterpart.

conditions play different roles: state separability restores a state-by-state expected-utility representation, while invariance removes the policy dependence of attention costs that would otherwise arise through the endogenous price distribution. We finally characterize optimal taxes depending on which of these conditions holds.

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). Relaxing state separability weakens, but does not overturn, the irrelevance logic. Consumers may then confuse different price realizations—e.g., they may misperceive a high price as a low one—and demand becomes interdependent across price realizations. But as long as attention costs are invariant, which is true with mutual information costs, no Pigouvian correction for inattention enters the optimal-tax formula. An appropriate version of our irrelevance result continues to apply.

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. To keep this possibility visible from the outset, the setup below allows consumers to process both producer prices and consumer prices. Under invariance, this richer description is immaterial. Once invariance fails, the distinction between the two prices lets the model speak to tax salience: the possibility that consumers respond differently to a tax change than to an otherwise equivalent posted-price change. We explore this possibility and its policy implications in Section 4. The present section instead focuses 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.[9]

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 after-tax (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 the other hand, is represented by the "cognitive" state $z = (\theta, \boldsymbol{p}, \boldsymbol{q})$, whose prior distribution is denoted by ϕ. At first glance, the distinction between z and θ may appear superfluous, because in equilibrium $\boldsymbol{p}$ and $\boldsymbol{q}$ will be deterministic transformations 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. Keeping both $\boldsymbol{p}$ and $\boldsymbol{q}$ in the cognitive state also lets us distinguish a tax change from an otherwise equivalent posted-price change. Equivalently, one can replace $\boldsymbol{q}$ by $\boldsymbol{\tau} = \boldsymbol{q} - \boldsymbol{p}$ and write the same cognitive state as $(\theta, \boldsymbol{p}, \boldsymbol{\tau})$. This alternative notation is closer to the language of tax salience, but the $(\boldsymbol{p}, \boldsymbol{q})$ representation is more convenient for the tax formulas below. By distinguishing z from θ, we can flexibly define the individual choice problem and then identify the condition on cognition that renders this richer description 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 absorbs 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,

[9] 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.

and the noise washes out on average.[10] Either way, the essence is that we abstract from 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.[11]

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

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

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 distribution of $a = (\boldsymbol{x}, y)$ conditional on $z = (\theta, \boldsymbol{p}, \boldsymbol{q})$, 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.

[10] To see what we 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 noise in total expenditure may vanish.

[11] 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 $\alpha : \Theta \times \mathbb{R}^N_+ \times \mathbb{R}^N_+ \to \mathbb{R}_{++}$ 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.[12]

Intuitively, state separability requires the cost of controlling mistakes in state z (e.g., when the price is high) to 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.[13] Finally, note 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).

[12] 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.

[13] 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.[14]

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 $g:\Theta\times\mathbb{R}^N_+\times\mathbb{R}^N_+\to\Theta\times\mathbb{R}^N_+\times\mathbb{R}^N_+$. We say that $(\tilde{F},\tilde{\phi})$ is the *transformation* of (F,ϕ) under g if:

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

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

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

[14] 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 $G \equiv \{g : \Theta \times \mathbb{R}^N_+ \times \mathbb{R}^N_+ \to \Theta \times \mathbb{R}^N_+ \times \mathbb{R}^N_+\}$ and fix any pair (F,ϕ). For any $g \in G$, let $(\tilde{F},\tilde{\phi})$ be the transformation of (F,ϕ) under g. The cost of attention is *invariant* if and only if, whenever $\tilde{\phi}$ is sufficient for ϕ with respect to F, we have $C(F,\phi) = C(\tilde{F},\tilde{\phi})$.

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 ($\boldsymbol{p},\boldsymbol{q}$), 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.[15]

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.[16] A concrete counterexample, on the other hand, is one where attention costs depend, not only on the mutual information between $a = (\boldsymbol{x}, y)$ and $z = (\theta, \boldsymbol{p}, \boldsymbol{q})$, 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.

[15] In fact, as this discussion suggests, we could relax our notion of invariance to apply only to a strict subset of G, given by transformations that preserve θ and relabel the price coordinates. That is, it is sufficient to have invariance over transformations of the producer- and consumer-price mappings.

[16] 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 represent the inattentive economy through a fictitious attentive economy, populated by agents with appropriately modified utilities. There are now H types of 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.[17] 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. For any consumer-price system $\boldsymbol{Q}$ and producer-price system $\boldsymbol{P}$, let $\phi^*_{\boldsymbol{Q},\boldsymbol{P}}$ denote the consistent prior: the prior that assigns probability π_θ to the cognition state $(\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$ for each θ. Throughout this section, until we introduce endogenous production, $\boldsymbol{P}$ is fixed, so we suppress it and write $\phi^*_{\boldsymbol{Q}}$ for $\phi^*_{\boldsymbol{Q},\boldsymbol{P}}$. 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 $\boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q},\boldsymbol{P}})$ and their expected utility, net of attention costs, is $V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q},\boldsymbol{P}})$, 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:

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

where $\beta^h > 0$ and $\lambda_\theta > 0$ represent, respectively, the Pareto weight of type-h agents and the shadow value of tax revenue in state θ, normalized so that the multiplier on revenue in state θ is $\pi_\theta \lambda_\theta$. 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 interdependencies, 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, for the prior

[17] The exact definitions of Φ and U^h are deferred to Appendix B.

$\phi^*_{\boldsymbol{Q},\boldsymbol{P}}$ induced by the price system, invariance lets us suppress the prior argument in the corresponding U^h: agents can "see through" different transformations of the underlying uncertainty. Second, if attention costs are also state separable, this object admits an expected utility representation: there exists a (possibly state-dependent) Bernoulli utility $\hat{u}^h : \Theta \times \mathscr{A} \to \mathbb{R}$ such that $U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q},\boldsymbol{P}}) = \sum_{\theta\in\Theta} \pi_\theta \hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta)$, for all $\boldsymbol{X}^h \equiv \left(\boldsymbol{x}^h_\theta\right)_{\theta\in\Theta}$, and all $\boldsymbol{Y}^h \equiv \left(y^h_\theta\right)_{\theta\in\Theta}$. By the same token, it is *as if* type-h agents solve a separate problem in each state, given by

$$\max_{\boldsymbol{x}^h_\theta, y^h_\theta} \left\{ \hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta) \text{ s.t. } \boldsymbol{q}_\theta \cdot \boldsymbol{x}^h_\theta + y^h_\theta \leq w^h_\theta \right\}. \tag{17}$$

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 Bernoulli utility $\hat{u}^h : \Theta \times \mathscr{A} \to \mathbb{R}$ such that, for all $(\boldsymbol{Q}, \boldsymbol{W}^h)$ and for $\phi = \phi^*_{\boldsymbol{Q},\boldsymbol{P}}$,*

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

where $\hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)$ and $\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)$ 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 after-tax prices $\boldsymbol{q}_\theta$—solve the following problem:

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

For each θ, this is mathematically identical to the Ramsey problems studied in Diamond and Mirrlees (1971b) and Diamond (1975), so we have essentially 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:

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

where $\partial_w v^h_\theta \equiv \frac{\partial \hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)}{\partial w}$ is the marginal *private value* of income in state θ and $\partial_w \bar{\boldsymbol{x}}^h_\theta \equiv \frac{\partial \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)}{\partial w}$ 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.[18] We can then state the following irrelevance result.

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

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

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 $\boldsymbol{S}(\boldsymbol{q}) \equiv \sum_{h \in \mathscr{H}} \mu^h \boldsymbol{S}^h(\boldsymbol{q})$ be, respectively, the aggregate demand function and the aggregate Slutsky matrix over the taxed goods; 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.

[18] 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.

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 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 insights 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:

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

where ν_i is the attention, or the precision, allocated to good i, $B_i(\cdot)$ is the corresponding benefit, and $C(\cdot)$ is the attention cost.[19] 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 $C(\nu_1,\nu_2)=(\nu_1+\nu_2)^k$ 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 i falls with q_i, and because now this is

[19] 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, $B_i(\nu_i,q_i)\equiv\max_{\bar{x}_i}\left(u_i(R_i(\nu_i)\bar{x}_i)-q_i\bar{x}_i\right)$, where $R_i(\nu_i)\equiv\left\{\mathbb{E}\left[(e^{\varepsilon_i})^{1-1/\eta_i}\right]\right\}^{1/(1-1/\eta_i)}$.

accompanied with higher expenditure on good $j \neq i$, the consumer may appear to be reallocating a fixed budget to the goods.

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 agent chooses consumption, she recognizes that her perception of after-tax 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: evaluated at the consistent prior, the indirect utility and demand functions of the fictitious attentive agents representing each group h, as defined in Section 3, can be written without the prior argument:

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q},\boldsymbol{P}}) = \tilde{v}^h(\boldsymbol{Q}, \boldsymbol{W}^h), \qquad \boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q},\boldsymbol{P}}) = \left(\tilde{\boldsymbol{x}}^h(\theta, \boldsymbol{Q}, \boldsymbol{W}^h)\right)_{\theta\in\Theta}.$$

The key 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 $\partial_{\boldsymbol{q}_\theta} \tilde{\boldsymbol{x}}^h_{\theta'} \equiv \left(\frac{\partial \tilde{x}^h_i(\theta', \boldsymbol{Q}, \boldsymbol{W}^h)}{\partial q_{\theta,j}}\right)_{i,j\in\mathscr{I}}$ denote the response of demand in state θ' to prices in state θ, and we similarly let $\partial_{w_\theta} \tilde{\boldsymbol{x}}^h_{\theta'} \equiv \frac{\partial \tilde{\boldsymbol{x}}^h(\theta', \boldsymbol{Q}, \boldsymbol{W}^h)}{\partial w_\theta}$ 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 $\boldsymbol{S}^h_{\theta,\theta'} = \mathbf{0}$ for all $\theta' \neq \theta$, we now allow $\boldsymbol{S}^h_{\theta,\theta'} \neq \mathbf{0}$.[20] We similarly adjust the marginal social value of income to account for any cross-state fiscal externalities:

$$\gamma^h_\theta \equiv \beta^h \partial_{w_\theta} \tilde{v}^h + \sum_{\theta' \in \Theta} \pi_{\theta'} \lambda_{\theta'} \boldsymbol{\tau}_{\theta'} \cdot \partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'}.$$

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*

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

Compared to Theorem 1, the important difference here is the inclusion of cross-state price elasticities, as captured in $\boldsymbol{S}^h_{\theta,\theta'}$ for $\theta' \neq \theta$. While these terms are generally non-zero, the planner still trades off revenue gains against deadweight losses, with no Pigouvian term for inattention itself. The trade-off is no longer separable across states, and the required sufficient statistics are correspondingly harder to measure, but the underlying Ramsey logic is unchanged. 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 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 spe-

[20] These cross-state Slutsky matrices need not be symmetric. Hence, the familiar textbook intuitions that rely on a symmetric Slutsky matrix do not automatically extend to this non-separable case.

cial 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 $\boldsymbol{\sigma} = (\sigma_k)_{k=0}^K \in \mathbb{R}^{K+1}$, 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 $k \in \{0,...,K\}$ 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*

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

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 after-tax 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 an untaxed quasi-linear numeraire. As in that section, the net utility of consuming x units of the taxed good 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, pre-tax, 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 after-tax 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 of q and the state dependence of after-tax prices. In this log-linear specification, φ governs the pass-through from producer prices to after-tax prices. 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 $\rho \equiv \frac{\mathbb{V}[\log q]}{\mathbb{V}[\varepsilon]+\mathbb{V}[\log q]} \in [0,1]$; this is the same as choosing the mutual information of ω and q. Finally, we let the agent's overall payoff be

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

where $K(\rho)$ is a strictly increasing and convex function, capturing the costs of attention.[21]

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 been 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 the optimal state-dependence parameter φ. To address this question, we first study how the agents' attention choice depends on the stochastic properties of the after-tax 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 after-tax prices are more volatile.*

Intuitively, choosing a ρ amounts to choosing the correlation between consumption and prices.

[21] 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 (47) 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.

For given $\rho < 1$, more volatile prices translate to 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 system leaves producer-price variation unchanged ($\varphi^* = 1$) if $\eta = 1$, dampens it ($\varphi^* < 1$) if $\eta < 1$, and amplifies it ($\varphi^* > 1$) 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 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 make after-tax prices more volatile, and thus stimulate attention, when $\eta > 1$, and to stabilize after-tax prices when $\eta < 1$. In both cases, the planner therefore aims at higher tax revenue; but the direction of the desired pass-through adjustment flips 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 dependence in taxes as an instrument for regulating attention (Proposition 4). At the same time, this example has shown that whether taxes should dampen or amplify producer-price variation 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 $\frac{\tau}{q} = \frac{\Lambda}{\eta}$.[22] 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.

[22] The relevant reduced-form elasticity here is η, not $\rho\eta$: when $\varphi = 1$, changing χ scales all after-tax prices but leaves $\mathbb{V}(\log q)$, and hence ρ, unchanged. We have therefore shut down the channel through which taxes affect attention, so the formula coincides with the fully attentive benchmark. This echoes Section 2.5: inattention does not necessarily lower demand elasticities; its effect depends on the policy margin and on how attention responds to it.

3.6 Adding production

We conclude this section by accommodating 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 D.

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. The feasible production set is convex and exhibits constant returns to scale. This technology is operated by a representative, fully-attentive, competitive firm. In each state, the firm thus maximizes profits, $\boldsymbol{p}_\theta \cdot \boldsymbol{x}^S + y^S$, subject to $H(\boldsymbol{x}^S, y^S, \theta) \leq 0$.[23] 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 in state θ, $G_\theta = \boldsymbol{p}_\theta \cdot \boldsymbol{g}_\theta$, must be financed by the same taxes as in our previous analysis.[24]

In this economy, both consumer (aftertax) and producer (pretax) prices are endogenous to policy, and the prior ϕ corresponds to the joint distribution of these two kinds of prices along with the underlying state of nature. In principle, this allows consumers to be differentially attentive to pre- and after-tax prices. Invariance, however, rules out this possibility and furthermore guarantees that consumers can "see through" both kinds of prices: their attention costs and their demands are invariant to mapping from θ to $(\boldsymbol{Q}, \boldsymbol{P})$. Along with the fact that the planner internalizes the consumers' attention costs, this ensures 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 ad-

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

[24] Note the government does not directly engage in production and that the utility of $\boldsymbol{g}$ is additively separable from the utility of the private goods are without serious loss of generality for our purposes. Both of these assumptions are without loss of generality for our purposes: government production can be replicated by market production by having the government hire a competitive firm to run its technology; and non-separable utility would let $\boldsymbol{g}$ enter the demand functions for the private goods and thus also tax revenue, without however changing our optimal tax formulas.

ditional conditions, which together characterize the entire solution to the planner's problem:[25]

$$\boldsymbol{p}_\theta = \frac{\nabla_{\boldsymbol{x}} H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)}{\nabla_y H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)},$$

$$\lambda_\theta \boldsymbol{p}_\theta = \sum_{h \in \mathscr{H}} \mu^h \beta^h \nabla_{\boldsymbol{g}} \tilde{u}^h(\boldsymbol{g}_\theta, \theta),$$

$$(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta) \cdot \sum_{h \in \mathscr{H}} \mu^h \boldsymbol{x}^h_\theta = \boldsymbol{p}_\theta \cdot \boldsymbol{g}_\theta, \qquad \begin{pmatrix} \boldsymbol{x}^S_\theta \\ y^S_\theta \end{pmatrix} = \sum_{h \in \mathscr{H}} \mu^h \begin{pmatrix} \boldsymbol{x}^h_\theta \\ y^h_\theta - w^h_\theta \end{pmatrix} + \begin{pmatrix} \boldsymbol{g}_\theta \\ 0 \end{pmatrix}.$$

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 these properties are intuitive, it is worth emphasizing that they rest on the assumption of invariant attention costs in two ways. The first was made explicit above: invariance guarantees that the planner can effectively ignore the endogeneity of ϕ^* when designing the optimal tax system. The second is implicit behind our assumption of a representative firm. By making this modeling choice, we have effectively presumed that production efficiency is attained with all firms facing the same set of prices (and thus also with no taxes on intermediate inputs). When firms are attentive, this property follows directly from the argument in Diamond and Mirrlees (1971a). When instead firms are inattentive, that argument extends provided the firms' attention costs are invariant: invariance guarantees that the planner cannot economize on firm attention costs by manipulating the price system. By contrast, production efficiency may fail if firm attention costs are non-invariant. This is the firm-side counterpart of the consumer-side logic we turn to next.

4 Relaxing invariance

We now identify the new margins that appear when attention costs fail to be invariant. First, violations of invariance open the door to cognitive externalities à la Angeletos and Sastry (2025). Attention costs may depend on the volatility, complexity, or other statistical properties of the price system, which consumers take as given but the government can regulate through its tax policy. Taxes can therefore serve not only the Ramsey function of the previous sections, but also a novel, Pigouvian function: simplifying the price system that agents have to process. We show that this translates to an additional rationale for state-dependent taxes, complementing that implied by non-separability. Second, relax-

[25] 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.

ing invariance brings tax salience into the analysis. With invariant attention costs, decomposing a consumer price into a producer price and a tax is only a relabeling. Once invariance fails, the same decomposition can affect attention costs and thereby demand, tax revenue and welfare. But as we will show, these effects could be placed "under the hood" of appropriately defined compensated elasticities. This section makes these two points explicit by deriving an optimal tax formula for a general class of economies with both endogenous production and arbitrary attention costs.

4.1 Cognitive externalities – a Pigouvian rationale for state-contingent taxes

In the preceding sections, invariance guaranteed that, although the planner could regulate ϕ^*, the equilibrium prior about prices (where "prices" now refer to both pretax and aftertax prices), this lever was inconsequential for both tax revenue and consumer surplus, and thus also for the design of optimal taxes. But now that attention costs are non-invariant, this lever becomes relevant.

Consider first how this matters for tax revenue. As in Theorem 2, the revenue effect of a consumer-price perturbation is summarized by appropriate Slutsky matrices. The only twist is that the compensated demand elasticities captured by these matrices are now *inclusive* of the indirect effect of taxes, or equivalently of the aftertax prices $\boldsymbol{Q}$, on demand via ϕ^*. Formally,

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

where $\nabla_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \left(\frac{dX^h_i(\theta',\boldsymbol{Q},\boldsymbol{W}^h,\phi^*)}{dq_{\theta,j}}\right)_{i,j\in\mathscr{J}}$ is now the matrix of *total* price derivatives, inclusive of the effect of $\boldsymbol{q}_\theta$ on $\bar{\boldsymbol{x}}^h_{\theta'}$ via ϕ^*, and $\partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \equiv \frac{\partial \boldsymbol{X}^h\left(\theta',\boldsymbol{Q},\boldsymbol{W}^h,\phi^*\right)}{\partial w_\theta}$ is the vector of income effects. Although these Slutsky matrices are different from those in Theorems 1 and 2, the common thread is that aftertax prices affect demand not only via classical substitution effects but also via attention choices. Importantly, these matrices are the relevant ones from a sufficient-statistics perspective: what matters for tax revenue are the overall compensated demand responses, not the precise mechanism behind them.

A more material difference emerges when we consider the effect of taxes on consumer surplus. This effect is now given by the sum of two terms:

$$\frac{d}{d\boldsymbol{q}_\theta} V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*) = -\left(\partial_{w_\theta} V^h\right)\bar{\boldsymbol{x}}^h_\theta - \boldsymbol{\Xi}^h_\theta, \qquad \text{with} \qquad \boldsymbol{\Xi}^h_\theta \equiv \nabla_{\boldsymbol{q}_\theta} C^h(F^h,\phi^*)\Big|_{F^h=F^{h,*}}.$$

The first term on the right-hand side of this equation 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. The second

term is new: it captures the relevant cognitive externality on type-h agents, namely the dependence of their attention costs on the complexity or the stochastic properties of the price system.[26]

We are now ready to state the following result, which shows the *exact* scope for Pigouvian taxation in the presence of rational—as opposed to behavioral—inattention.

Theorem 3. *Consider the general environment of Section 3.1, where production is (possibly) endogenous, now allowing for general attention costs (possibly neither state-separable nor invariant). The optimal taxes satisfy*

$$\sum_{h\in\mathscr{H}} \mu^h \left[\left(\pi_\theta \lambda_\theta - \gamma_\theta^h\right) \bar{\boldsymbol{x}}_\theta^h + \sum_{\theta'\in\Theta} \pi_{\theta'} \lambda_{\theta'} \left(\boldsymbol{S}_{\theta,\theta'}^h\right)^\top \boldsymbol{\tau}_{\theta'} \right] = \pi_\theta \lambda_\theta \Xi_\theta \qquad \forall \theta, \tag{27}$$

where $\Xi_\theta \equiv \frac{1}{\lambda_\theta \pi_\theta} \sum_{h\in\mathscr{H}} \mu^h \beta^h \Xi_\theta^h$ *measures the aggregate cognitive externality in state* θ *(normalized by that state's shadow value).*

Relative to Theorems 1 and 2, the optimal taxes now contain a Pigouvian correction, equal to the Pareto-weighted average cognitive externality. As anticipated, this new term requires a violation of invariance and thus a departure from mutual-information costs. Furthermore, and unlike what is suggested by some of the behavioral public finance literature, the microfoundation, sign, and size of this term is, at least in principle, orthogonal to the question of whether inattention makes demand less or more responsive to taxes. Instead, this term reflects exclusively how attention costs depend on the statistical properties of the price system, which consumers take as given but the government can regulate through its tax policy. Finally, Theorem 3 is at once deceptively simple and deeply informative: it puts no restriction on the degree, scope and form of inattention, other than it being rational. Notably, by allowing for arbitrary attention costs, the theorem also allows for differential attention to pre- and after-tax prices; it follows that our optimal tax formula applies regardless of the degree of "tax salience" and any empirical evidence about it.

We elaborate on the richness, the precise meaning, and the practical implications of Theorem 3 in the rest of this section. We first illustrate the Pigouvian wedge Ξ_θ, which as just shown emerges only away from the benchmark of invariant (or mutual-information) costs. We then zero in on how exactly tax salience matters, or does not matter, under the rational-inattention framework.

[26] Here $\nabla_{\boldsymbol{q}_\theta}$ denotes the total equilibrium derivative, including the induced adjustment of producer prices. Thus Ξ_θ^h incorporates the dependence of ϕ^* on the tax system, but holds the attention rule F^h fixed at $F^{h,*}$. This isolates the externality—which is relevant for Pigouvian purposes—from the behavioral response.

4.2 Cognitive externalities – a Pigouvian rationale for state-dependent taxes

To illustrate the Pigouvian wedge Ξ_θ of non-invariant attention costs, in this section we momentarily abstract from endogenous production and conduct two exercises. First, we translate Theorem 3 to settings in which the cognitive externality takes a flexible parametric form—with attention costs depending on the stochastic properties of the price system via a finite set of moments. Second, we illustrate the result in the case of a single taxable good.

State-dependent taxes as an instrument for easing attention costs

Echoing Corollaries 1 and 2, we abstract from distributional concerns and from any uncertainty other than in prices. We furthermore suppose 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. With the same abuse of notation as in those corollaries, we re-express demand as $\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma})$ and attention costs as $C^h(F^h,\phi) = C^h(F^h,\boldsymbol{\sigma})$. To keep notation compact, we finally write $\Psi_k \equiv \sum_{h\in\mathcal{H}} \mu^h \beta^h \partial_{\sigma_k} C^h(F^h,\boldsymbol{\sigma})$ for the aggregate cognitive wedge associated with moment k, and we represent the latter as $\sigma_k = m_k(\boldsymbol{q})$. We can then recast the optimal tax formula (27) as follows.

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

$$\boldsymbol{S}(\boldsymbol{q},\boldsymbol{\sigma})^\top \boldsymbol{\tau} + \sum_k \mathbb{E}\big[\boldsymbol{\tau}\cdot\partial_{\sigma_k}\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma})\big]\partial_{\boldsymbol{q}} m_k(\boldsymbol{q}) = -\Lambda\hat{\boldsymbol{x}}(\boldsymbol{q},\boldsymbol{\sigma}) + (1-\Lambda)\sum_k \Psi_k \partial_{\boldsymbol{q}} m_k(\boldsymbol{q}). \tag{28}$$

The left-hand side of this tax formula captures the total effect of changing $\boldsymbol{\tau}$ on tax revenue. Echoing Corollary 2, this encompasses—via the second term—the effect of taxes on the moments of the price distribution and thereby on the optimal attention choices and the associated demands. Turning to the right-hand side, the first term captures the standard effect on consumer surplus. What's new is the very last term, which captures the cognitive externalities. In particular, the wedges Ψ_k seen above are the translations of the wedges Ξ_θ seen in Theorem 3, now re-expressed in terms of the *moments* of the tax system (or equivalently of the aftertax prices). Altogether, the optimal taxes therefore combine the Ramsey function highlighted in Corollary 2, with the new, Pigou-like function implied by the violation of invariance. But the practical takeaway is similar in the following sense: non-separability and non-invariance provide complementary rationales for manipulating the stochastic properties of

the price system and thus for state-dependent taxes.

We further illustrate this complementary rationale for state-dependent taxes below, using a variant of the example considered in Section 3.5. In the original example, we had assumed mutual-information costs, guaranteeing that attention costs were invariant to the stochastic properties of the price system. Here, we relax this property in order to capture a scenario in which agents make relatively larger mistakes when they live in a more volatile, or more complex, market environment.

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 rationale for state-dependent taxes due to non-separability (already characterized in Corollary 2 and Proposition 4) and isolate instead the new, complementary, rationale due to 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. This lets us capture, within the log-normal setting of this example, the possible dependence of attention costs on the "sparsity", the "simplicity", or the "transparency" of the price system. Intuitively, a consumer who is confused by more volatile prices is effectively one who prefers a more sparse price system; and conversely, a scenario where attention costs decrease when prices are more volatile can be interpreted as one where consumers can more easily evaluate the costs and benefits of their choices across states when those states are further apart from one another in terms of their associated prices.

To illustrate these ideas and motivate the cost function we will accommodate in the present example, consider first an example in which attention costs are commensurate to the 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 squared correlation between ω and $\log q$. 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 this 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 a cognitive externality. A cognitive externality instead emerges whenever $K_v \neq 0$. The case of $K_v > 0$ captures the idea in the above example: the cognitive burden of tracking prices and choosing the right consumption is higher when prices are more volatile (or less sparse). The opposite case of $K_v < 0$ could capture a situation in which consumers have an *easier* time discerning, and optimally responding to, price changes when these changes are larger (and thus more transparent). Let $\rho(v)$ denote the optimal attention choice as a function of after-tax price volatility, and let $\epsilon(v) \equiv \frac{\partial \log \rho(v)}{\partial \log v}$ be the corresponding elasticity. To simplify the exposition, the upcoming proposition assumes that, around the optimum, $\rho(v)$ is positive and differentiable and does not fall too sharply with price volatility: $\epsilon(v) > -1/2$.[27]

As in Section 3.5, the planner uses a tax system so that $q = \chi p^{\varphi}$, where $\chi > 0$ parameterizes the level of taxes and $\varphi > 0$ parameterizes their state-dependence; in particular, the planner can regulate v, the volatility of prices, by choosing $\varphi \neq 1$. In that section, the only rationale for manipulating v was to influence attention choices in the direction of raising more tax revenue. Here, this particular rationale has been switched off by restricting $\eta = 1$, but a new rationale emerges because of the violation of invariance and the resulting cognitive externality. The next result shows how the planner uses state dependence in taxes to ease that externality.

Proposition 6. *Consider the economy described above, where attention costs can be expressed as $K(\rho, v)$. Suppose the optimal state-dependence parameter φ^* is interior and that, at the optimum, $\rho(v)$ is positive and differentiable, with $\epsilon(v) \equiv \frac{\partial \log \rho(v)}{\partial \log v} > -1/2$. The optimal tax system dampens the variation in aftertax prices if $K_v > 0$ and amplifies it 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 $K_v(\rho(v), v)$ is the relevant cognitive externality.

Proposition 6 shows that non-invariant costs, just like non-separable costs, can introduce state-

[27] The condition is immediate if the constraint $\rho \leq 1$ binds around the optimum, so that $\rho(v) = 1$. If attention is interior and $K_{\rho\rho} > 0$, differentiating the attention FOC $K_\rho(\rho, v) = v/2$ shows that a sufficient condition is $\rho K_{\rho\rho} + v - 2vK_{\rho v} > 0$ at the optimum. For example, $K_{\rho v} \leq 1/2$ is enough.

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 $K_\upsilon > 0$, this goal is achieved by reducing the volatility in after-tax prices, which in turn is achieved by dampening the pass-through from producer prices. When $K_\upsilon < 0$, the same logic calls for increasing after-tax volatility, and hence for amplifying that pass-through. Proposition 6 further shows that the absolute *magnitude* of the optimal state dependence 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 especially simple here because supply is infinitely elastic (pre-tax 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 with a visual illustration. In particular, set

$$K(\rho, \upsilon) = \frac{\kappa}{2}\left(\mathbb{V}[\log q] - \mathbb{E}\left[\mathbb{V}[\log q \mid \omega]\right]\right)^2 = \frac{\kappa}{2}(\rho\,\upsilon)^2$$

for $\kappa > 0$, as in the example from the beginning of this section; but now allow $\eta \neq 1$, and plot, 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 dampens producer-price variation ($\varphi^* < 1$), reflecting the same force: $K_\upsilon > 0$.[28] 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 directions. In all cases, the magnitude of the state-dependence increases with κ.

4.3 Tax salience

We now zero in on how tax salience matters, or does not matter, for optimal taxation within a rational-inattention framework. The irrelevance results of Theorems 1 and 2 ruled out a meaningful role for tax salience by assuming invariant attention costs. This assumption made sure that, regardless of how inattentive consumers were to taxes and prices, their choices depended only on the effective, aftertax, prices—the planner could not manipulate their choices by decomposing the same consumer price to

[28] For the $\eta = 1$ curve, the quadratic variance-reduction specification satisfies the regularity condition in Proposition 6 over the plotted range. The unconstrained attention choice is $\rho^u(\upsilon) = 1/(2\kappa\upsilon)$. At the relevant optimum this exceeds one, so that $\rho(\upsilon) = 1$ and hence $\epsilon(\upsilon) = 0$. See Appendix C.3 for details.

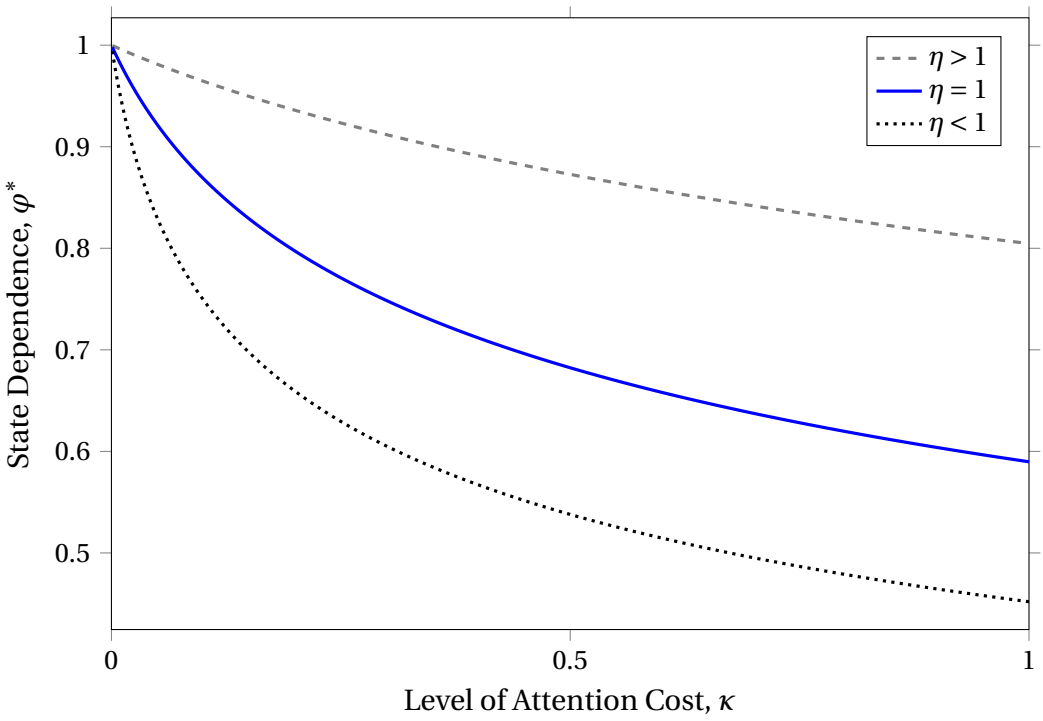


Figure 1: Optimal state dependence φ^* with non-invariant costs.

different combinations of producer prices and taxes. Non-invariance is thus *necessary* for tax salience to play a meaningful role, but it is not *sufficient*: with exogenous producer prices, any given consumer price maps to a unique combination of producer prices and taxes, so the aforementioned lever is just absent. It follows that a meaningful discussion of tax salience within the rational-inattention framework requires, not only a violation of invariance, but also an endogenous response of producer prices to taxes—an insight that has no analogue in the related behavioral public finance literature.

In this subsection, we expand on this insight and on its policy implications. In particular, we unpack the price elasticities (Slutsky matrices) featured in Theorem 3 and show how exactly tax salience enters them when, and only when, attention costs are non-invariant and production is endogenous. We then explain why the planner does not have to look "under the hood" and can instead use directly the tax formula of Theorem 3, provided that she properly identifies the relevant elasticities via a tax reform. And since the only substantive assumption behind this theorem is that inattention is rational (which in turn also implies that a benevolent planner ought to internalize people's attention costs), the upshot here is this: salience can affect behavior and measured elasticities, but it does not create an additional behavioral-welfare correction.

An instructive example The simplest way to see the role of tax salience is to revisit the economy of Section 2, where there was one taxed good, its utility was the same in all states, and attention costs were separable across states, but now make two modifications: we let supply be imperfectly elastic, so that producer prices adjust in response to taxes; and we let consumers pay differential attention to the tax and to the pretax price, due to a failure of invariance.[29] The latter modification amounts to letting demand, in each state, be a function of both the consumer and the producer price in the same state:

$$\bar{x} = \bar{x}(q, p).$$

(To ease notation, in this example, just as in Section 2, we suppress the dependence of $\bar{x}, q$ and p and any other endogenous object on θ.) As already explained, the emergence of p as a separate determinant of consumer demand is rationalized by a violation of invariance. Concretely, note that demand could be expressed equivalently as a function of the producer price and the consumer tax: $\bar{x} = \tilde{x}(\tau, p)$, with $\tilde{x}(\tau, p) \equiv \bar{x}(p + \tau, p)$. Taxes and pretax prices are equally salient if and only if $\partial\bar{x}/\partial p = \partial\tilde{x}/\partial\tau$ for all (τ, p), or equivalently $\partial\bar{x}/\partial p = 0$ for all (q, p). Taxes are instead less salient if and only if $-\partial\bar{x}/\partial p > -\partial\tilde{x}/\partial\tau$ for all (τ, p), or equivalently $\partial\bar{x}/\partial p < 0$ for all (q, p). Intuitively, if taxes are less salient than pretax prices, demand can be increased by decomposing the same consumer price into a higher tax and a lower producer price.

The production side is summarized by an inverse supply function $p = S(x)$, with $S'(x) > 0$. Market clearing then imposes that, in equilibrium, the producer and the consumer price jointly satisfy

$$p = S\big(\bar{x}(q, p)\big).$$

We assume that the market-clearing condition can be solved locally for p as a differentiable function P of q.[30] This means that we can still treat the consumer price q as the effective policy instrument—with the corresponding producer price and tax being determined in the background as $p = P(q)$ and $\tau = T(q) \equiv q - P(q)$. The optimal policy is then given by the following specialization of Theorem 3.

[29] Concretely, we let attention costs be

$$C(F, \phi) = \sum_{(q,p)} \phi(q, p)\mu(q, p) \int_{\mathscr{X}} K(f(x|q, p))dx, \tag{29}$$

for some exogenous function μ. This cost specification preserves separability, as in Section 2, but relaxes invariance: the marginal cost of attention is allowed to vary with q and p via the "weighting" function μ.

[30] By the implicit-function theorem, a sufficient condition is that $1 - S'(\bar{x}(q, p))\partial\bar{x}(q, p)/\partial p \neq 0$. In the less-salient-tax case described above, $\partial\bar{x}(q, p)/\partial p < 0$; since $S' > 0$, this condition is automatically satisfied.

Proposition 7. *In this one-good, state-separable economy, the optimal tax satisfies*

$$\frac{\tau}{q} = \frac{1}{\epsilon^{GE}}\left(\Lambda - \frac{\Xi}{\bar{x}}\right), \tag{30}$$

where $\epsilon_\theta^{GE} \equiv -\frac{d\log \bar{x}(q,P(q))}{d\log q}$ *is the elasticity of equilibrium consumption with respect to a change in the consumer price,* Λ *captures the social value of tax revenue, and* Ξ *captures the Pigouvian correction (if any).*

Compared to the tax formula obtained in the original version of our one-good example (see Proposition 1), the one seen above contains two differences. The first is the Pigouvian correction highlighted in Theorem 3. The second is a subtle change in the price elasticity that enters the tax formula. Here, we put aside the Pigouvian correction and zero in on the second change.

Before, demand was only a function of the aftertax price, so the elasticity of the former with respect to the latter was unambiguously defined. Here, we can instead define two distinct elasticities: a partial or "micro" elasticity, which holds the producer price p fixed as we vary the consumer price q; and a total or "GE" elasticity, which accounts for how the producer price adjusts in equilibrium as the planner varies q (by varying the tax). Clearly, these two elasticities are different if and only if (i) supply is imperfectly elastic and (ii) producer prices enter demand separately from consumer prices. Our result then shows that, whenever this is true, the "right" sufficient statistic is the total or GE elasticity.

To see more clearly what this elasticity contains, and how tax salience interacts with the production side in shaping the difference between it and the more plain-vanilla, partial, elasticity, evaluate the following objects at the optimal allocation:

$$\epsilon^q \equiv -\frac{\partial \log \bar{x}(q,p)}{\partial \log q}, \qquad \epsilon^p \equiv -\frac{\partial \log \bar{x}(q,p)}{\partial \log p}, \qquad \eta \equiv \frac{dS^{-1}(p)}{dp}\frac{p}{S^{-1}(p)}.$$

Here ϵ^q is the consumer-price elasticity holding the producer price fixed, ϵ^p is the elasticity with respect to the producer price holding the consumer price fixed (which, recall, is non-zero only if taxes and pretax prices have different salience), and η is the supply elasticity. Demand gives $d\log \bar{x} = -\epsilon^q d\log q - \epsilon^p d\log p$, while market clearing along the supply curve gives $d\log p = \eta^{-1} d\log \bar{x}$. Together, these equations describe the incidence feedback: a higher consumer price lowers demand directly; market clearing then lowers the producer price; and, when demand also responds to producer prices, this induced price movement feeds back into demand. Solving for the total response

gives the GE elasticity, which shows up in the optimal tax formula (30), as

$$\epsilon^{GE} = \frac{\epsilon^q}{1+\epsilon^p/\eta}, \tag{31}$$

and the associated pass-through of consumer prices to producer prices as

$$\frac{dp}{dq} = -\frac{p}{q}\frac{\epsilon^{GE}}{\eta}. \tag{32}$$

If attention costs are invariant, $\epsilon^p_\theta = 0$. If supply is perfectly elastic, $\eta = \infty$. In either case, the feedback through producer prices drops out and the distinction between the partial and GE elasticities vanishes ($\epsilon^{GE} = \epsilon^q$). For tax salience to enter the Ramsey tradeoff, it is thus necessary that we depart from *both* of these cases: invariance must fail to let producer prices enter demand and, in addition, supply must be imperfectly elastic to let producer prices change with taxes.

This discussion also clarifies the empirical counterpart of the tax formula. The relevant elasticity is the response of quantity to a tax reform, normalized by the realized change in the consumer price after markets have settled. In math:

$$\epsilon^{GE} = -\frac{d\log\bar{x}/d\log\tau}{d\log q/d\log\tau}.$$

Put differently, one can estimate ϵ^{GE} by a simple 2SLS procedure: use an exogenous tax reform as an instrument in a regression of $\bar{x}$ on q. Importantly, one should *not* try to control for p in this regression: the relevant sufficient statistic contains the endogenous response of producer prices to the tax reform. Controlling for p would help identify ϵ^q, but what matters for the Ramsey tradeoff is ϵ^{GE}, not ϵ^q.

This is also where the connection to behavioral public finance becomes transparent. In Chetty et al. (2009), the comparison is between a tax added at the register and an equivalent change in the posted price. In our notation, the register experiment varies q holding p fixed and identifies ϵ^q. The posted-price experiment moves p and q one-for-one, holding the statutory tax fixed; expressed as an elasticity with respect to q, this experiment identifies $\epsilon^q + (q/p)\epsilon^p$. The gap between the two experiments therefore recovers the salience elasticity. Given a supply elasticity, these objects assemble into the elasticity in (31). In this sense, salience changes which experiment recovers the elasticity that belongs in the tax formula. But if the planner uses a tax reform to identify ϵ^{GE} directly, the planner can sidestep all these other experiments and conduct "business as usual". From this perspective, the

available evidence on tax salience does *not* change how the Ramsey trade off should be resolved.

What remains is the reinterpretation of Chetty et al. (2009) as evidence of non-invariance, which in turn opens the door to a Pigouvian correction. This correction, however, cannot be pinned down by that evidence. To quantify Ξ, one must know how attention costs vary with the price system—and while the evidence Chetty et al. (2009) offers some clues, it does not identify the attention costs that may rationalize the observed behaviors. The identification of attention costs is indeed an empirical challenge more generally. Caplin and Dean (2015) have made important progress by asking when attention costs can be recovered by observed choice in an abstract decision-theoretic setting, but the bridge from their analysis to public finance and our results remains to be built. Till such a bridge is built or Ξ is otherwise quantified, it is sensible to abstract from cognitive externalities and focus on the Ramsey trade off. The practical takeaway from Proposition 7 is then that the planner should just estimate the relevant GE elasticity and not worry about the underlying cognitive mechanisms.

Translation to general setting Appendix XX verifies that the same logic extends directly to the full environment with many goods, many states, heterogeneous consumers, income effects, and general attention costs. Theorem 3 is deliberately stated in terms of total equilibrium responses: the Slutsky matrices appearing there are the general counterparts of the elasticity ϵ^{GE} in Proposition 7. As in equation (31), each response combines the direct effect of changing consumer prices while holding producer prices fixed with the producer-price feedback generated by market clearing. This decomposition clarifies how tax salience enters the theorem, but it is not necessary for optimal tax design: the relevant sufficient statistics are the total equilibrium responses appearing in the theorem, regardless of the cognitive mechanisms and production feedbacks behind them.

5 Discussion

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

Main lessons: 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 behavior; 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 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 translated to a two-part rationale for state-dependence in the optimal taxes. In the one case, state-dependence helped relax the trade-off between collecting tax revenue and distorting consumption; in the other, it helped ease people's cognitive burden.

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 across multiple cognitive states); and second, their attention costs 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 how 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 which consumers preferred "sparsity" or "simplicity": consumers may still economize attention costs by pooling their demands across multiple price realizations, but they no longer 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 from that found in Farhi and Gabaix (2020): the planner aims, not at undoing the effects of inattention, but rather at coarsening the price system or otherwise easing people's cognitive burden.

6 Conclusion

This paper studied optimal commodity taxation when consumers are rationally inattentive. Our central finding was an irrelevance result: policy makers should neither try to undo the effects of inattention nor reconsider the classical sufficient-statistics formula for optimal taxes; instead, they can design taxes *as if* agents were fully attentive. We establish the sharpest version of this result under two restrictions on attention costs—invariance and separability. We then showed how violations of these two conditions provided complementary rationales for state-dependent taxes, but also pointed out that the available evidence does not yet speak to the quantitative importance of these insights. We further showed that the possible role of tax salience can be subsumed behind the appropriate equilibrium elasticities.

These findings qualify the normative arguments found in an emerging behavioral public finance

literature. Behaviors akin to sparsity, narrow bracketing, and mental accounting are consistent not only with rational inattention but also 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 further showed that a complementary, non-Pigouvian, rationale for state-dependent taxes emerged when attention costs were invariant but non-separable, the case pertinent to mutual-information costs. In this case, state-dependent taxes can ease the Ramsey tradeoff by regulating the optimal attention choice. Both of these insights are too speculative at this point to be of practical value, but they deserve 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 extent 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 Proofs for Section 2

We prove Proposition 1 and Proposition 2. The first result establishes the inverse-elasticity formula under the separable attention cost (3). The second result shows how endogenous attention can raise the effective elasticity of demand and thereby lower optimal taxes.

A.1 Proof of Proposition 1

If attention costs are given by (3), the conditional density f_q can be chosen separately for each realized aftertax price q. For any target $\bar{x}$, define the fictitious Bernoulli utility

$$\hat{u}(\bar{x}) \equiv \max_{f \geq 0} \left\{ \int_{\mathscr{X}} \big(u(x) f(x) - K(f(x))\big)\, dx \quad \text{s.t.} \quad \int_{\mathscr{X}} f(x)\, dx = 1, \quad \int_{\mathscr{X}} x f(x)\, dx = \bar{x} \right\},$$

and let

$$\hat{v}(q) \equiv \max_{\bar{x}} \{\hat{u}(\bar{x}) - q\bar{x}\}, \qquad \hat{x}(q) \equiv \arg\max_{\bar{x}} \{\hat{u}(\bar{x}) - q\bar{x}\}.$$

Because the objective in the definition of $\hat{u}$ is concave in f and the constraints are linear, $\hat{u}$ is concave in the target $\bar{x}$. It follows that $\bar{x} \mapsto \hat{u}(\bar{x}) - q\bar{x}$ is concave, so the usual monotonicity of demand applies. At the equilibrium prior $\phi^*_{\boldsymbol{Q}}$, the probability of a realized price q is $\phi^*_{\boldsymbol{Q}}(q) = \sum_{\{\theta : q_\theta = q\}} \pi_\theta$. Hence

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

Similarly, aggregate demand in state θ is

$$\bar{x}(q_\theta, \phi^*_{\boldsymbol{Q}}) = \hat{x}(q_\theta).$$

These identities do not require the mapping $\theta \mapsto q_\theta$ to be one-to-one: if several states generate the same aftertax price, they all use the same conditional problem and their probabilities are simply aggregated in $\phi^*_{\boldsymbol{Q}}(q)$.

The Ramsey problem (2) therefore reduces to

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

The first-order condition with respect to q_θ is

$$\pi_\theta \frac{\partial \hat{v}(q_\theta)}{\partial q_\theta} + \lambda \pi_\theta \left\{ \hat{x}(q_\theta) + (q_\theta - p_\theta) \frac{\partial \hat{x}(q_\theta)}{\partial q_\theta} \right\} = 0. \tag{33}$$

Using Roy's identity on the fictitious statewise problem, $\frac{\partial \hat{v}(q_\theta)}{\partial q_\theta} = -\hat{x}(q_\theta)$. Substituting into (33) and simplifying by dividing out $\pi_\theta > 0$:

$$-\hat{x}(q_\theta) + \lambda \left\{ \hat{x}(q_\theta) + \tau_\theta \frac{\partial \hat{x}(q_\theta)}{\partial q_\theta} \right\} = 0.$$

Dividing by $\lambda \hat{x}(q_\theta)$, rearranging, and using the elasticity definition $\epsilon(q_\theta) \equiv -\frac{\partial \log \hat{x}(q_\theta)}{\partial \log q_\theta}$:

$$1 - \frac{\tau_\theta}{q_\theta} \epsilon(q_\theta) = \frac{1}{\lambda}.$$

Solving for $\frac{\tau_\theta}{q_\theta}$ and using $\Lambda \equiv 1 - \frac{1}{\lambda}$, we arrive at equation (4):

$$\frac{\tau_\theta}{q_\theta} = \frac{\Lambda}{\epsilon(q_\theta)}.$$

A.2 Proof of Proposition 2

Under the CRRA and log-normal specifications in the main text, the risk-adjusted return evaluates to $R(v) = \exp\left\{-\frac{1}{2\eta v}\right\}$. The optimal target from the inner problem is

$$x^\circ(q, v) = q^{-\eta} R(v)^{\eta - 1} = \exp\left\{-\eta \log q - \tfrac{\eta - 1}{2\eta} v^{-1}\right\}. \tag{34}$$

The portfolio-choice interpretation developed in the main text clarifies the comparative statics. A higher v translates to lower risk and thus a higher risk-adjusted return $R(v)$; but a higher return has competing income and substitution effects on the optimal "investment" x°, so the overall impact of v on x° is ambiguous:

$$\frac{\partial \log x^\circ(q, v)}{\partial v} = \frac{\eta - 1}{2\eta} v^{-2}.$$

When $\eta > 1$, substitution effects dominate and $\frac{\partial x^\circ}{\partial v} > 0$: the agent targets higher consumption when attention is higher. When $\eta < 1$, income effects dominate and the sign reverses.

Holding attention fixed, the elasticity of demand with respect to q is η, just as in the frictionless benchmark; but the *optimal* attention varies with q, and this changes the overall elasticity. For $\eta \neq 1$,

the benefit function evaluated at the optimal target is

$$B(\nu, q) \equiv u(R(\nu)x^\circ(q,\nu)) - qx^\circ(q,\nu) = \frac{q\,x^\circ(q,\nu) - \eta}{\eta - 1}.$$

By the envelope theorem, $\frac{\partial B}{\partial q} = -x^\circ$ and therefore $\frac{\partial^2 B}{\partial q \partial \nu} = -\frac{\partial x^\circ}{\partial \nu}$. At an interior optimum, $\nu^\circ(q)$ solves $\frac{\partial B}{\partial \nu} = C'(\nu)$, and the implicit function theorem gives

$$\frac{d\nu^\circ}{dq} = -\frac{\frac{\partial^2 B}{\partial q \partial \nu}}{\frac{\partial^2 B}{\partial \nu^2} - C''(\nu)} = \frac{\frac{\partial x^\circ}{\partial \nu}}{\frac{\partial^2 B}{\partial \nu^2} - C''(\nu)}.$$

The second-order condition makes the denominator negative, so the sign of $\frac{d\nu^\circ}{dq}$ is opposite to the sign of $\frac{\partial x^\circ}{\partial \nu}$. For $\eta > 1$, a higher q reduces the marginal benefit of attention ($\frac{\partial^2 B}{\partial q \partial \nu} < 0$), lowering its optimal level ($\frac{d\nu^\circ}{dq} < 0$). Since lower attention also discourages consumption, a higher q reduces demand both directly and through reduced attention.

When $\eta < 1$, the sign of $\frac{\partial x^\circ}{\partial \nu}$ flips: more attention now *reduces* the optimal target. But then the signs of $\frac{\partial^2 B}{\partial q \partial \nu}$ and $\frac{d\nu^\circ}{dq}$ also flip, so the product $\frac{\partial \log x^\circ}{\partial \nu}\frac{d\nu^\circ}{d\log q}$ retains the same sign. A higher q still discourages demand via its effect on ν, and the elasticity is again higher than η. In the knife-edge case $\eta = 1$, $\frac{\partial x^\circ}{\partial \nu} = 0$ and $\frac{d\nu^\circ}{dq} = 0$, so the two elasticities coincide.

Collecting these observations, the effective elasticity of demand is

$$\epsilon(q) = -\frac{d\log x^\circ(q,\nu^\circ(q))}{d\log q} = \eta - \frac{\partial \log x^\circ}{\partial \nu}\frac{d\nu^\circ}{d\log q} \;\geq\; \eta,$$

with strict inequality at any interior optimum whenever $\eta \neq 1$. If $\eta = 1$, $x^\circ(q,\nu) = q^{-1}$ is independent of ν, so attention does not change the elasticity and $\epsilon(q) = \eta = 1$. By Proposition 1, the optimal tax is therefore weakly lower than its frictionless counterpart, and strictly lower at interior optima when $\eta \neq 1$.

A.2.1 Epstein-Zin preferences

To disentangle risk aversion from the substitution parameter η, suppose the agent evaluates lotteries using a CRRA certainty equivalent with coefficient $\gamma > 0$, with the log case understood as the usual limit:

$$\mathrm{CE}_\gamma(x) = \left(\mathbb{E}[x^{1-\gamma}]\right)^{1/(1-\gamma)}, \tag{35}$$

while utility over the certainty-equivalent level retains the same isoelastic form u from the main text, with elasticity η. As before, let realized consumption be $x = \bar{x}\, e^{\varepsilon}$ with $\varepsilon \sim \mathcal{N}(-\frac{1}{2\nu}, \nu^{-1})$, so that $\mathbb{E}[e^{\varepsilon}] = 1$. The certainty equivalent evaluates to

$$\mathrm{CE}_{\gamma}(x) = \bar{x}\, e^{-\gamma/(2\nu)}. \tag{36}$$

Maximizing $u(\mathrm{CE}_{\gamma}(x)) - q\,\bar{x}$ over $\bar{x}$ yields the FOC

$$\bar{x}^{-1/\eta}\, e^{-\gamma(1-1/\eta)/(2\nu)} = q. \tag{37}$$

Solving, the optimal target $x^{\circ}(q,\nu)$ satisfies

$$\log x^{\circ}(q,\nu) = -\eta \log q + \frac{\gamma(1-\eta)}{2\nu}. \tag{38}$$

The sign of the effect of attention on x° depends only on η, confirming that it is the substitution parameter, not risk aversion, that governs how attention affects demand:

- If $\eta > 1$, substitution effects dominate, and x° *increases* with ν.
- If $\eta = 1$, x° does not depend on ν.
- If $\eta < 1$, income effects dominate, and x° *decreases* with ν.

B Proofs for Section 3

We first construct the fictitious attentive economy used in Section 3. Lemma 1 establishes the equivalence with the original inattentive economy. Lemma 2 shows that, under invariance and state separability, indirect utility and demand decompose state by state. We then prove Theorem 1. Finally, we prove Theorem 2, which keeps invariance but relaxes state separability.

B.1 Auxiliary lemmas for the general-model representation

We give the formal details behind the auxiliary lemmas used in Section 3. Fix a type h and a producer-price system $\boldsymbol{P} \equiv (\boldsymbol{p}_{\theta})_{\theta\in\Theta}$. Let Φ denote the set of priors over cognition states whose support is spanned by θ. Formally, for each $\theta \in \Theta$ there is a unique cognition state $z_{\phi}(\theta) = (\theta, \boldsymbol{p}_{\theta}, \boldsymbol{q}_{\theta})$ in the

support of ϕ, and ϕ assigns positive probability only to $\{z_\phi(\theta) : \theta \in \Theta\}$. This set is defined without reference to equilibrium; it contains, but is not limited to, consistent priors. We let Z_ϕ denote the support of this prior.

For a given consumer-price system $\boldsymbol{Q} \equiv (\boldsymbol{q}_\theta)_{\theta\in\Theta}$, we say that ϕ is *consistent* with $(\boldsymbol{Q}, \boldsymbol{P})$ if $z_\phi(\theta) = (\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$ and $\phi(z_\phi(\theta)) = \pi_\theta$ for all θ, and write $\phi^*_{\boldsymbol{Q}} \in \Phi$ for this unique consistent prior, suppressing the fixed $\boldsymbol{P}$ in the notation. Let $\mathscr{A} \subset \mathbb{R}^{N+1}_+$ denote the space of consumption bundles $(\boldsymbol{x}, y)$, where $\boldsymbol{x} \in \mathbb{R}^N_+$ collects the N taxable goods and $y \in \mathbb{R}_+$ is the non-taxable numeraire. Write $\boldsymbol{X}^h \equiv (\boldsymbol{x}^h_\theta)_{\theta\in\Theta}$ and $\boldsymbol{Y}^h \equiv (y^h_\theta)_{\theta\in\Theta}$ and define the modified utility object:

$$\begin{aligned} U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi) \equiv \max_{F^h \in \mathscr{F}} \Big\{ \sum_{\theta\in\Theta} \pi_\theta \int_{\mathscr{A}} u^h(\boldsymbol{x}, y, \theta)\, dF^h(\boldsymbol{x}, y \mid z_\phi(\theta)) - C^h(F^h, \phi) \\ \text{s.t.} \int_{\mathscr{A}} (\boldsymbol{x}, y)\, dF^h(\boldsymbol{x}, y \mid z_\phi(\theta)) = (\boldsymbol{x}^h_\theta, y^h_\theta) \quad \forall \theta \in \Theta \Big\}. \end{aligned} \tag{39}$$

Given $(\boldsymbol{Q}, \boldsymbol{W}^h, \phi)$, define the fictitious fully-attentive consumer problem:

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi) \equiv \max_{\boldsymbol{X}^h, \boldsymbol{Y}^h} \Big\{ U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi) \text{ s.t. } \boldsymbol{q}_\theta \cdot \boldsymbol{x}^h_\theta + y^h_\theta \le w^h_\theta \ \forall \theta \in \Theta \Big\}, \tag{40}$$

and let $\boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi)$ denote the associated maximizer for the taxable goods. We assume throughout that the maxima in (39), (40), in the original type-h inattentive problem (12) of Section 3, and in the statewise problem (17) are attained.

Lemma 1. *For each h, there exists a function U^h such that, in equilibrium, the average demand of the type-h consumers in our inattentive economy is given by $\boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}})$ and their expected utility, net of attention costs, is given by $V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}})$.*

Proof. Fix h, $\boldsymbol{Q}, \boldsymbol{W}^h$, and a consistent prior $\phi = \phi^*_{\boldsymbol{Q}}$. Let $F^{h,*}$ solve the original type-h inattentive problem (12). Because ϕ is consistent, the objective of (12) is exactly (39) evaluated at $F^{h,*}$. By linearity of expectations, feasibility of $F^{h,*}$ in the original problem implies that its induced conditional means $(\bar{\boldsymbol{X}}^h, \bar{\boldsymbol{Y}}^h) \equiv (\bar{\boldsymbol{x}}^h_\theta, \bar{y}^h_\theta)_{\theta\in\Theta}$ satisfy $\boldsymbol{q}_\theta \cdot \bar{\boldsymbol{x}}^h_\theta + \bar{y}^h_\theta \le w^h_\theta$ for all θ. Therefore $(\bar{\boldsymbol{X}}^h, \bar{\boldsymbol{Y}}^h)$ is feasible in (40). Moreover, $F^{h,*}$ is feasible in the definition (39) of $U^h(\bar{\boldsymbol{X}}^h, \bar{\boldsymbol{Y}}^h, \phi)$, hence

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi) \ge U^h(\bar{\boldsymbol{X}}^h, \bar{\boldsymbol{Y}}^h, \phi) \ge \text{value of (12)}.$$

Conversely, let $(\boldsymbol{X}^h, \boldsymbol{Y}^h)$ be feasible in (40), and let $F^h_{X,Y}$ solve (39) for $(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi)$. The budget con-

straint of (40) ensures that $F^h_{X,Y}$ is affordable on average, so it is feasible in (12). Therefore the value of (12) is at least $U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi)$ for every feasible $(\boldsymbol{X}^h, \boldsymbol{Y}^h)$, and thus at least $V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi)$. Combining both inequalities, the two problems have the same value and $(\bar{\boldsymbol{X}}^h, \bar{\boldsymbol{Y}}^h)$ is a maximizer in (40). Evaluating at $\phi = \phi^*_{\boldsymbol{Q}}$ yields the stated equilibrium mapping. □

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

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

where $\hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)$ and $\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)$ are, respectively, the max *and the* argmax *of problem (17) from the main text.*

Proof. Under state separability (Definition 1), for any $\phi \in \Phi$ the attention cost takes the form

$$C^h(F^h, \phi) = \sum_{\theta \in \Theta} \alpha^h(z_\phi(\theta)) \, \phi(z_\phi(\theta)) \int_{\mathscr{A}} K^h(f^h(\boldsymbol{x}, y \mid z_\phi(\theta))) \, d\boldsymbol{x} \, dy,$$

where $\alpha^h : \Theta \times \mathbb{R}^N_+ \times \mathbb{R}^N_+ \to \mathbb{R}_{++}$ is the weighting function from (13). Since each $\phi \in \Phi$ assigns a unique cognition state $z_\phi(\theta) = (\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$ to each θ, the cost is additively separable across states. A priori, however, the weight $\alpha^h(z_\phi(\theta)) = \alpha^h(\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$ may depend on the producer and after-tax prices and thus on the tax policy.

Invariance (Definition 4) forces the weighting function α^h to be independent of the price components. Take any $\phi \in \Phi$ and any $F^h \in \mathscr{F}$. Fix any alternative price systems $(\boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)_{\theta \in \Theta}$ and consider the transformation g defined by $g(\theta, \boldsymbol{p}, \boldsymbol{q}) = (\theta, \boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)$ for all $(\boldsymbol{p}, \boldsymbol{q})$, which relabels the price components of each state θ. Let $(\tilde{F}^h, \tilde{\phi})$ be the transformation of (F^h, ϕ) under g. Recall that the support of ϕ is $Z_\phi = \{z_\phi(\theta) : \theta \in \Theta\}$ where $z_\phi(\theta) = (\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$. Since g preserves θ, it is one-to-one on Z_ϕ. By construction, $\tilde{\phi}((\theta, \boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)) = \phi(z_\phi(\theta))$ and $\tilde{f}^h(\boldsymbol{x}, y \mid (\theta, \boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)) = f^h(\boldsymbol{x}, y \mid z_\phi(\theta))$ for each θ. Hence $\tilde{\phi}$ is sufficient for ϕ with respect to F^h, and invariance gives $C^h(F^h, \phi) = C^h(\tilde{F}^h, \tilde{\phi})$. Expanding both sides using separability:

$$\sum_{\theta \in \Theta} \alpha^h(\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta) \, \phi(z_\phi(\theta)) \int_{\mathscr{A}} K^h(f^h) \, d\boldsymbol{x} \, dy = \sum_{\theta \in \Theta} \alpha^h(\theta, \boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta) \, \phi(z_\phi(\theta)) \int_{\mathscr{A}} K^h(f^h) \, d\boldsymbol{x} \, dy.$$

Because F^h is arbitrary, we can freely vary the conditional action densities $f^h(\cdot \mid z_\phi(\theta))$ across different states θ. For the equality to hold for all such variations, the coefficients must match state-by-state,

which requires $\alpha^h(\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta) = \alpha^h(\theta, \boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)$ for each θ. As the alternative price systems $(\boldsymbol{p}'_\theta, \boldsymbol{q}'_\theta)_{\theta\in\Theta}$ were arbitrary, we can define $\tilde{\alpha}^h(\theta) \equiv \alpha^h(\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$, which depends only on the exogenous state θ.

Now specialize to the consistent prior $\phi = \phi^*_{\boldsymbol{Q}}$, for which $\phi(z_\phi(\theta)) = \pi_\theta$ for all θ. Since $\tilde{\alpha}^h(\theta)$ depends only on the physical state, substituting into (39) gives

$$U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q}}) = \max_{F^h\in\mathscr{F}} \left\{ \sum_{\theta\in\Theta} \pi_\theta \left[\int_{\mathscr{A}} u^h(\boldsymbol{x}, y, \theta)\, dF^h(\boldsymbol{x}, y \mid z_\phi(\theta)) - \tilde{\alpha}^h(\theta) \int_{\mathscr{A}} K^h(f^h(\boldsymbol{x}, y \mid z_\phi(\theta)))\, d\boldsymbol{x}\, dy \right] \right\},$$

subject to

$$\int_{\mathscr{A}} (\boldsymbol{x}, y)\, dF^h(\boldsymbol{x}, y \mid z_\phi(\theta)) = (\boldsymbol{x}^h_\theta,\, y^h_\theta) \qquad \forall \theta \in \Theta.$$

The objective is additively separable across states, and the constraints also apply state by state. Therefore,

$$U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q}}) = \sum_{\theta\in\Theta} \pi_\theta\, \hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta),$$

where $\hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta)$ is the value of the problem in state θ:

$$\hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta) \equiv \max_{f\geq 0} \left\{ \int_{\mathscr{A}} \Big(u^h(\boldsymbol{x}, y, \theta)\, f(\boldsymbol{x}, y \mid \theta) - \tilde{\alpha}^h(\theta)\, K^h(f(\boldsymbol{x}, y \mid \theta)) \Big)\, d\boldsymbol{x}\, dy \right\}$$

subject to

$$\int_{\mathscr{A}} (\boldsymbol{x}, y)\, f(\boldsymbol{x}, y \mid \theta)\, d\boldsymbol{x}\, dy = (\boldsymbol{x}^h_\theta,\, y^h_\theta) \quad \text{and} \quad \int_{\mathscr{A}} f(\boldsymbol{x}, y \mid \theta)\, d\boldsymbol{x}\, dy = 1.$$

Because $\tilde{\alpha}^h(\theta)$ depends only on θ, the function $\hat{u}^h$ is independent of $\boldsymbol{Q}$. Substituting into (40):

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}}) = \max_{(\boldsymbol{x}^h_\theta, y^h_\theta)_{\theta\in\Theta}} \left\{ \sum_{\theta\in\Theta} \pi_\theta\, \hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta) \text{ s.t. } \boldsymbol{q}_\theta \cdot \boldsymbol{x}^h_\theta + y^h_\theta \leq w^h_\theta \ \forall \theta \in \Theta \right\}.$$

Since the objective is additively separable and the budget constraints apply state-by-state, the maximization problem decomposes into independent statewise subproblems:

$$\hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta) \equiv \max_{\boldsymbol{x}^h_\theta, y^h_\theta} \left\{ \hat{u}^h(\theta, \boldsymbol{x}^h_\theta, y^h_\theta) \text{ s.t. } \boldsymbol{q}_\theta \cdot \boldsymbol{x}^h_\theta + y^h_\theta \leq w^h_\theta \right\},$$

which is exactly (17) from the main text. It follows that

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

B.2 Proof of Theorem 1

The proof starts from the general Ramsey problem and uses the fictitious attentive representation constructed above. For every h, let $v^h(\phi)$ and $F^h(\phi)$ be, respectively, the value and the optimal stochastic choice rule of the original inattentive agent's problem given consistent prior $\phi \in \Phi$. Next, let $\bar{\boldsymbol{x}}^h_\theta(\phi) \equiv \int_{\mathscr{A}} \boldsymbol{x}\, dF^h(\boldsymbol{x}, y \mid z_\phi(\theta); \phi)$ be the average demand generated by $F^h(\phi)$ in state θ. Let $\boldsymbol{Q} \equiv \left(\boldsymbol{q}_\theta\right)_{\theta\in\Theta}$ collect the after-tax prices across all goods and states, where $\boldsymbol{q}_\theta = \boldsymbol{p}_\theta + \boldsymbol{\tau}_\theta$. Finally, let $\phi^*_{\boldsymbol{Q}}$ denote the unique prior consistent with $\boldsymbol{Q}$. We can then write the Ramsey problem for our general economy as:

$$\max_{\boldsymbol{Q}} \sum_{h\in\mathscr{H}} \mu^h \beta^h v^h(\phi^*_{\boldsymbol{Q}}) + \sum_{\theta\in\Theta} \pi_\theta \lambda_\theta \left(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\right) \cdot \sum_{h\in\mathscr{H}} \mu^h \bar{\boldsymbol{x}}^h_\theta(\phi^*_{\boldsymbol{Q}}). \tag{41}$$

By Lemma 1, in equilibrium the agent's indirect utility and the aggregate demands can be replaced with the indirect utility function V^h and the aggregate demand function $\boldsymbol{X}^h$ of the fictitious attentive economy. That is, $v^h(\phi^*_{\boldsymbol{Q}}) = V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}})$ and $\bar{\boldsymbol{x}}^h_\theta(\phi^*_{\boldsymbol{Q}}) = \boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}})$. Moreover, under invariance and state separability, when evaluated at the consistent prior $\phi^*_{\boldsymbol{Q}}$, the indirect utility and aggregate demand mappings further admit a state-by-state decomposition by Lemma 2:

$$V^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}}) = \sum_{\theta\in\Theta} \pi_\theta \hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta) \quad \text{and} \quad \boldsymbol{X}^h(\boldsymbol{Q}, \boldsymbol{W}^h, \phi^*_{\boldsymbol{Q}}) = \left(\hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)\right)_{\theta\in\Theta}.$$

Hence, the Ramsey problem (41) reduces to a collection of standard state-by-state Ramsey problems. That is, the optimal after-tax prices $\boldsymbol{q}_\theta$ in state θ solve:

$$\max_{\boldsymbol{q}_\theta} \left\{ \sum_{h\in\mathscr{H}} \mu^h \beta^h \hat{v}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta) + \lambda_\theta \left(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\right) \cdot \sum_{h\in\mathscr{H}} \mu^h \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta) \right\}. \tag{42}$$

The first-order condition with respect to $\boldsymbol{q}_\theta$ is the vector equation

$$\sum_{h\in\mathscr{H}} \mu^h \beta^h \frac{\partial \hat{v}^h}{\partial \boldsymbol{q}_\theta} + \lambda_\theta \sum_{h\in\mathscr{H}} \mu^h \left[\bar{\boldsymbol{x}}^h_\theta + \left(\partial_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_\theta\right)^\top \boldsymbol{\tau}_\theta\right] = \boldsymbol{0},$$

where $\bar{\boldsymbol{x}}^h_\theta \equiv \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)$ is the aggregate demand of type h in state θ and $\partial_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_\theta \equiv \frac{\partial \hat{\boldsymbol{x}}^h(\theta, \boldsymbol{q}_\theta, w^h_\theta)}{\partial \boldsymbol{q}_\theta}$. Roy's identity gives $\frac{\partial \hat{v}^h}{\partial \boldsymbol{q}_\theta} = -(\partial_w \hat{v}^h_\theta)\, \bar{\boldsymbol{x}}^h_\theta$, where $\partial_w \hat{v}^h_\theta \equiv \frac{\partial \hat{v}^h_\theta}{\partial w}$ is the marginal private value of income in state θ. Rearranging, we obtain

$$\sum_{h\in\mathscr{H}} \mu^h \left[(\lambda_\theta - \beta^h \partial_w \hat{v}^h_\theta)\, \bar{\boldsymbol{x}}^h_\theta + \lambda_\theta \left(\partial_{\boldsymbol{q}_\theta} \bar{\boldsymbol{x}}^h_\theta\right)^\top \boldsymbol{\tau}_\theta\right] = \boldsymbol{0}.$$

Now, using the definition of the Slutsky matrix

$$\boldsymbol{S}_\theta^h \equiv \partial_{\boldsymbol{q}_\theta} \tilde{\boldsymbol{x}}_\theta^h + (\partial_w \tilde{\boldsymbol{x}}_\theta^h)(\tilde{\boldsymbol{x}}_\theta^h)^\top,$$

we have $(\partial_{\boldsymbol{q}_\theta} \tilde{\boldsymbol{x}}_\theta^h)^\top \boldsymbol{\tau}_\theta = (\boldsymbol{S}_\theta^h)^\top \boldsymbol{\tau}_\theta - (\boldsymbol{\tau}_\theta \cdot \partial_w \tilde{\boldsymbol{x}}_\theta^h)\tilde{\boldsymbol{x}}_\theta^h$. Defining $\gamma_\theta^h \equiv \beta^h \partial_w \hat{v}_\theta^h + \lambda_\theta \boldsymbol{\tau}_\theta \cdot \partial_w \tilde{\boldsymbol{x}}_\theta^h$, we arrive at

$$\sum_{h \in \mathscr{H}} \mu^h \left[(\lambda_\theta - \gamma_\theta^h)\, \tilde{\boldsymbol{x}}_\theta^h + \lambda_\theta \boldsymbol{S}_\theta^h \boldsymbol{\tau}_\theta \right] = \boldsymbol{0},$$

which matches (20) in the main text.[31]

B.3 Proof of Theorem 2

We first show that with invariance the aggregate demand and indirect utility of the fictitious economy, evaluated at a consistent prior, can be written without the prior argument ϕ. Fix a consumer-price system $\boldsymbol{Q}$, let $\phi = \phi^*_{\boldsymbol{Q}}$ denote the corresponding consistent prior, and take any feasible $F^h \in \mathscr{F}$. Fix any alternative consumer-price system $\boldsymbol{Q}'$, let $\phi' = \phi^*_{\boldsymbol{Q}'}$, and consider the relabeling g that maps $z_\phi(\theta) = (\theta, \boldsymbol{p}_\theta, \boldsymbol{q}_\theta)$ into $z_{\phi'}(\theta) = (\theta, \boldsymbol{p}_\theta, \boldsymbol{q}'_\theta)$ for each $\theta \in \Theta$. Since g preserves the physical state, it is one-to-one on Z_ϕ. Because both priors are consistent with the same producer-price system, the transformed prior is exactly ϕ'. Let $(\tilde{F}^h, \phi')$ be the transformation of (F^h, ϕ) under g. By construction,

$$\tilde{f}^h(\boldsymbol{x}, y \mid z_{\phi'}(\theta)) = f^h(\boldsymbol{x}, y \mid z_\phi(\theta)) \qquad \forall \theta \in \Theta,$$

so ϕ' is sufficient for ϕ with respect to F^h. Invariance therefore implies $C^h(F^h, \phi) = C^h(\tilde{F}^h, \phi')$.

The same relabeling leaves unchanged the expected-utility term in (39) and the state-by-state mean constraints. Because both priors are consistent, the expected-utility term is evaluated with the same weights π_θ. Hence every feasible $F^h \in \mathscr{F}$ has exactly the same value under ϕ and under ϕ', which yields

$$U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q}}) = U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q}'})$$

for every alternative price system $\boldsymbol{Q}'$. Therefore there exists a function $\tilde{U}^h$ such that

$$U^h(\boldsymbol{X}^h, \boldsymbol{Y}^h, \phi^*_{\boldsymbol{Q}}) = \tilde{U}^h(\boldsymbol{X}^h, \boldsymbol{Y}^h)$$

[31] The intermediate FOC naturally gives $(\boldsymbol{S}_\theta^h)^\top \boldsymbol{\tau}_\theta$. Under invariance and state separability, $\boldsymbol{S}_\theta^h$ is symmetric by Lemma 2, so this equals $\boldsymbol{S}_\theta^h \boldsymbol{\tau}_\theta$, as written in Theorem 1.

for every consumer-price system $\boldsymbol{Q}$. Thus invariance removes dependence on the consumer-price component of the cognition states once we restrict attention to consistent priors with fixed $\boldsymbol{P}$. Substituting $\tilde{U}^h$ into (40), we can write

$$V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) = \tilde{v}^h(\boldsymbol{Q},\boldsymbol{W}^h) \qquad \text{and} \qquad \boldsymbol{X}^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) = \left(\tilde{\boldsymbol{x}}^h(\theta,\boldsymbol{Q},\boldsymbol{W}^h)\right)_{\theta\in\Theta}.$$

Thus, in the nonseparable case, the relevant Ramsey problem can be written as:

$$\max_{\boldsymbol{Q}} \sum_{h\in\mathscr{H}} \mu^h \beta^h \tilde{v}^h(\boldsymbol{Q},\boldsymbol{W}^h) + \sum_{\theta\in\Theta} \pi_\theta \lambda_\theta \left(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\right) \cdot \sum_{h\in\mathscr{H}} \mu^h \tilde{\boldsymbol{x}}^h(\theta,\boldsymbol{Q},\boldsymbol{W}^h), \tag{43}$$

where the multiplier on revenue in state θ is $\pi_\theta\lambda_\theta$, as in the main text. Unlike the state-separable case, the indirect utility $\tilde{v}^h$ and demand $\tilde{\boldsymbol{x}}^h(\theta,\cdot)$ depend on the entire price vector $\boldsymbol{Q}$, so the problem does not decompose state by state. Nonetheless, the prior ϕ has been eliminated: the planner's first-order conditions involve only the standard Roy and Slutsky channels.

Using Roy's identity and the notation from Section 3.4, the first-order condition of (43) with respect to the price of good i in state θ can be written as

$$\sum_{h\in\mathscr{H}} \mu^h \left[-\beta^h (\partial_{w_\theta} \tilde{v}^h) \bar{x}^h_{i,\theta} + \pi_\theta \lambda_\theta \bar{x}^h_{i,\theta} + \sum_{\theta'\in\Theta} \pi_{\theta'} \lambda_{\theta'} \sum_{n=1}^{N} \tau_{n,\theta'} \frac{\partial \bar{x}^h_{n,\theta'}}{\partial q_{i,\theta}} \right] = 0. \tag{44}$$

The key difference from the separable case is that $\frac{\partial \bar{x}^h_{n,\theta'}}{\partial q_{i,\theta}}$ need not vanish for $\theta' \neq \theta$.

Rearranging the definition of the cross-state Slutsky matrix from Section 3.4,

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

we can write the uncompensated derivative as $\frac{\partial \bar{x}^h_{n,\theta'}}{\partial q_{i,\theta}} = \left(\boldsymbol{S}^h_{\theta,\theta'}\right)_{ni} - \partial_{w_\theta} \bar{x}^h_{n,\theta'} \bar{x}^h_{i,\theta}$. Substituting into (44), summing over n, and collecting income-effect terms yields:

$$\sum_{h\in\mathscr{H}} \mu^h \left[\left\{ \pi_\theta\lambda_\theta - \beta^h \partial_{w_\theta} \tilde{v}^h - \sum_{\theta'\in\Theta} \pi_{\theta'}\lambda_{\theta'} \boldsymbol{\tau}_{\theta'} \cdot \partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'} \right\} \bar{\boldsymbol{x}}^h_\theta + \sum_{\theta'\in\Theta} \pi_{\theta'}\lambda_{\theta'} \left(\boldsymbol{S}^h_{\theta,\theta'}\right)^\top \boldsymbol{\tau}_{\theta'} \right] = \boldsymbol{0}. \tag{45}$$

Using the definition of γ^h_θ from Section 3.4, together with $V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) = \tilde{v}^h(\boldsymbol{Q},\boldsymbol{W}^h)$, we arrive at the

vector equation reported in Theorem 2:

$$\sum_{h\in\mathscr{H}} \mu^h \left[(\pi_\theta \lambda_\theta - \gamma_\theta^h)\tilde{\boldsymbol{x}}_\theta^h + \sum_{\theta'\in\Theta} \pi_{\theta'}\lambda_{\theta'} \left(\boldsymbol{S}_{\theta,\theta'}^h\right)^\top \boldsymbol{\tau}_{\theta'} \right] = \mathbf{0}.$$

C Proofs for the Log-Normal Examples

We derive the three log-normal results used in Sections 3.5 and 4.2. The first two propositions use mutual-information costs. The final proposition allows the attention cost to depend directly on after-tax price volatility through a general cost $K(\rho, v)$; the variance-reduction specification discussed in the main text is one special case. Throughout, let $\gamma \equiv 1/\eta$ denote the CRRA curvature parameter, so that the demand elasticity in the main text satisfies $\eta = 1/\gamma$.

Let the after-tax price be $q(\theta) = \chi\theta^\varphi$. Under log-normal fundamentals $\log\theta \sim \mathcal{N}(0, v_\theta)$, we have $\log q \sim \mathcal{N}(\log\chi, v)$, where $v \equiv \varphi^2 v_\theta$ is the variance of the after-tax log price. As in the main text, the signal is $\omega = \log q + \varepsilon$, yielding the signal-to-noise ratio $\rho \equiv \frac{v}{v+\mathbb{V}[\varepsilon]} \in [0,1]$.

The posterior mean of $\log q$ is linear in the signal, and the posterior variance is $(1-\rho)v$. Hence the optimal individual demand is the one in (25). Let $\bar{X}(p;\chi,\varphi) \equiv \mathbb{E}[X(\omega) \mid p]$ denote average demand conditional on the producer price p and the tax parameters (χ,φ). We suppress the dependence on (χ,φ) when no confusion arises. Integrating over the distribution of q, one can analytically evaluate the expected utility benefits of attention, defined as $b(\rho;\chi,v) \equiv \mathbb{E}[\frac{X^{1-\gamma}}{1-\gamma}] - \mathbb{E}[qX]$. This yields:

$$b(\rho;\chi,v) = \frac{\gamma}{1-\gamma}\exp\left\{\frac{\gamma-1}{\gamma}\log\chi + \frac{\gamma-1}{2\gamma^2}(\gamma-\rho)v\right\}. \tag{46}$$

C.1 Proof of Proposition 3

With multiplicative mutual-information costs, the agent's objective is to choose ρ to maximize

$$b(\rho;\chi,v)\exp\{-(1-\gamma)K(\rho)\},$$

where $K(\rho)$ is strictly convex. Dropping terms that are constant in ρ, this is equivalent to maximizing

$$\frac{v}{2\gamma^2}\rho - K(\rho).$$

The first-order condition with respect to ρ is:

$$K'(\rho) = \frac{v}{2\gamma^2} = \frac{\varphi^2 v_\theta}{2\gamma^2}. \tag{47}$$

Equivalently, $K'(\rho) = \frac{1}{2}\eta^2 v$. Because $K(\cdot)$ is strictly convex, K' is strictly increasing. It follows that the optimal attention choice is weakly increasing in the after-tax price volatility v, and strictly increasing whenever the optimum is interior.

C.2 Proof of Proposition 4

After substituting the optimal attention rule $\rho(\varphi)$, the relevant part of the planner's Lagrangian can be written as

$$A(\varphi)\, b\big(\rho(\varphi); \chi, \varphi^2 v_\theta\big) + \lambda\, \mathbb{E}\big[(q(p) - p)\bar{X}(p; \chi, \varphi)\big],$$

where λ denotes the Lagrange multiplier on the revenue constraint and $A(\varphi) \equiv \exp\{-(1-\gamma)K(\rho(\varphi))\}$ is the attention-cost scaling factor. As in Section 2, we focus on the case in which the shadow value of public funds is high enough for positive taxes to be relevant. With the multiplicative attention-cost term, this condition is $\lambda > A(\varphi)$ at the relevant optimum. Write the associated fiscal wedge as

$$\Omega(\varphi) \;\equiv\; 1 - \frac{A(\varphi)}{\lambda}. \tag{48}$$

Thus Ω measures how much the shadow cost of public funds exceeds the private utility cost of attention. Using the envelope theorem for ρ and suppressing the arguments of $\bar{X}$, the first-order condition for χ reduces (after dividing by λ) to

$$\mathbb{E}[p\bar{X}] \;=\; \big(1 - \gamma\,\Omega(\varphi)\big)\,\mathbb{E}[q\bar{X}]. \tag{49}$$

Because both expectations are positive, (49) implies $1 - \gamma\,\Omega(\varphi) > 0$ at the optimum. Again using the envelope theorem for ρ, differentiate the Lagrangian with respect to φ, substitute (49), and evaluate the resulting log-normal moments to obtain:

$$\big(1-\varphi\big)\Big[\big(1-\gamma\,\Omega(\varphi)\big)\,\rho(\varphi) \;+\; \varphi\,\frac{d\rho}{d\varphi}\Big] \;-\; \varphi\,\frac{d\rho}{d\varphi}\,\Omega(\varphi)\Big(\gamma \;-\; \tfrac{\gamma+1}{2}\,\varphi\Big) \;=\; 0. \tag{50}$$

To sign $1-\varphi^*$, use the identity $\gamma - \frac{\gamma+1}{2}\varphi = \frac{\gamma-1}{2} + \frac{\gamma+1}{2}(1-\varphi)$ to rewrite the last term in (50) and collect

all factors of $(1-\varphi)$ on the left-hand side:

$$(1-\varphi)\left[(1-\gamma\Omega)\rho + \left(1-\Omega\tfrac{\gamma+1}{2}\right)\varphi\frac{d\rho}{d\varphi}\right] = \varphi\frac{d\rho}{d\varphi}\Omega\frac{\gamma-1}{2}. \tag{51}$$

The term in square brackets on the left-hand side is strictly positive. To see this, note:

1. $1-\gamma\Omega(\varphi) > 0$ by (49), and $\rho > 0$ at an interior optimum.
2. $\Omega(\varphi)\frac{\gamma+1}{2} < 1$. If $\gamma < 1$, then $\frac{\gamma+1}{2} < 1$ and $\Omega < 1$, so the product is below 1. If $\gamma > 1$, then $1-\gamma\Omega > 0$ implies $\Omega < 1/\gamma$, so $\Omega\frac{\gamma+1}{2} < \frac{\gamma+1}{2\gamma} < 1$.

Since the left-hand bracket is positive and, in the revenue-raising case, $\varphi\, d\rho/d\varphi\,\Omega > 0$, the sign of $(1-\varphi)$ equals the sign of $(\gamma-1)$. Translating to $\eta = 1/\gamma$:

- If $\eta = 1$ ($\gamma = 1$), the right-hand side is zero, so $\varphi^* = 1$. The tax system is state-independent.
- If $\eta < 1$ ($\gamma > 1$), then $\varphi^* < 1$: the after-tax price is concave in θ.
- If $\eta > 1$ ($\gamma < 1$), then $\varphi^* > 1$: the after-tax price is convex in θ.

C.3 Proof of Proposition 6

Suppose attention costs depend on both mutual information and after-tax price volatility, so the cost is $K(\rho, v)$. Holding (χ, φ) fixed, the part of the agent's objective that depends on ρ is

$$\frac{v}{2\gamma^2}\rho - K(\rho, v).$$

Let $\rho(v)$ denote the induced optimal attention rule. When the optimal attention choice is interior, it satisfies

$$K_\rho(\rho(v), v) = \frac{v}{2\gamma^2},$$

or equivalently $K_\rho(\rho(v), v) = \frac{1}{2}\eta^2 v$. If instead the constraint $\rho \le 1$ binds strictly, $K_\rho(1, v) < v/(2\gamma^2)$, then the constrained optimum is $\rho(v) = 1$ around that point and $\rho'(v) = 0$. The argument below only requires that the constrained attention choice be positive and differentiable at the optimum.

The planner's problem now includes the direct cognitive externality of price volatility $v = \varphi^2 v_\theta$ on attention costs. In addition to the standard revenue and substitution channels, there is now an extra term involving $K_v(\rho, v)$, the partial derivative of the cost with respect to volatility. Specializing the

planner's first-order condition to the log-utility limit $\gamma \to 1$ (equivalently $\eta \to 1$, which shuts down the state-dependence motive from Proposition 4), using $\frac{dv}{d\varphi} = 2\varphi v_\theta$, and dividing by $v_\theta > 0$, we obtain

$$(1-\varphi)\left[\rho(v) + \varphi \frac{d\rho}{d\varphi}\right] - 2\varphi K_v(\rho(v), v) = 0. \tag{52}$$

Rearranging and evaluating at the optimum, we obtain:

$$1 - \varphi^* = \frac{2\varphi^* K_v(\rho(v), v)}{\rho(v) + \varphi^* \frac{d\rho}{d\varphi}}. \tag{53}$$

Since $v = \varphi^2 v_\theta$, we have $\frac{dv}{d\varphi} = 2\varphi v_\theta$, and therefore $\varphi\, d\rho/d\varphi = 2v\,\rho'(v)$. The denominator becomes

$$\rho(v) + \varphi \frac{d\rho}{d\varphi} = \rho(v) + 2v\,\rho'(v) = \rho(v)\big(1 + 2\epsilon(v)\big),$$

where $\epsilon(v) \equiv \rho'(v)\,v/\rho(v)$ is the elasticity of optimal attention with respect to price volatility as defined in the main text. Substituting,

$$1 - \varphi^* = \left.\frac{2\varphi^* K_v(\rho(v), v)}{\rho(v)\big(1 + 2\epsilon(v)\big)}\right|_{v=(\varphi^*)^2 v_\theta}. \tag{54}$$

Since $1 + 2\epsilon(v) > 0$ by assumption, $\text{sign}(1-\varphi^*) = \text{sign}(K_v)$. Hence, when higher price volatility directly increases attention costs ($K_v > 0$), the planner lowers the pass-through from producer prices to after-tax prices, implying $1 - \varphi^* > 0$ and thus $\varphi^* < 1$. When $K_v < 0$, the sign reverses and $\varphi^* > 1$.

For the $\eta = 1$ curve in Figure 1, specialize the cost to the quadratic variance-reduction form used in the main text: $K(\rho, v) = \kappa(\rho v)^2/2$. The unconstrained attention FOC then gives $\rho^u(v) = 1/(2\kappa v)$, so the constraint $\rho \le 1$ binds around the relevant optimum whenever $\kappa v^* < 1/2$, with $v^* = (\varphi^*)^2 v_\theta$. When this condition holds, the planner's first-order condition reduces to

$$1 - \varphi^* = 2\kappa v_\theta (\varphi^*)^3.$$

It remains to check when this candidate is consistent with the attention constraint binding. The first-order condition implies

$$2\kappa v^* = 2\kappa v_\theta (\varphi^*)^2 = \frac{1-\varphi^*}{\varphi^*},$$

so $\kappa v^* < 1/2$ is equivalent to $\varphi^* > 1/2$. To express this in terms of primitives, let $a \equiv \kappa v_\theta$ and define

$H(\varphi) \equiv 1 - \varphi - 2a\varphi^3$. The function H is strictly decreasing, with $H(0) > 0$ and $H(1) < 0$, so the binding first-order condition has a unique solution, and this solution lies in $(0,1)$. It lies above $1/2$ iff $H(1/2) > 0$, or equivalently $a < 2$. Thus $\kappa v_\theta < 2$ is sufficient for the attention constraint to bind around the optimum. The plotted range has $v_\theta = 1$ and $\kappa \leq 1$, so the constraint binds around the optimum and $\epsilon(v) = 0$.

D Endogenous Production

This appendix completes the endogenous-production discussion from the main text. We first prove Proposition 5. Under invariance, endogenous production does not change the consumer-tax formulas from Theorems 1 and 2. Production only adds the conditions that determine producer prices and public good provision. We then prove the general non-invariance theorem and unpack the demand responses and cognitive-wedges that enter the formula. We conclude by proving Proposition 7 and note how firm-side inattention could lead to genuine departures from production efficiency.

D.1 Proof of Proposition 5

We maintain the general setup from the main text and endogenize production. In each state θ, a competitive representative firm produces the taxable goods $\boldsymbol{x}^S_\theta \in \mathbb{R}^N_+$ and net output of the numeraire $y^S_\theta \in \mathbb{R}$ subject to a technology constraint $H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta) \leq 0$. The feasible production set is convex and exhibits constant returns to scale, and H is differentiable with positive gradient on the relevant frontier. Profit maximization implies

$$\boldsymbol{p}_\theta = \frac{\nabla_{\boldsymbol{x}} H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)}{\nabla_y H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)}. \tag{55}$$

Constant returns to scale also imply zero profits at the supporting prices,

$$\boldsymbol{p}_\theta \cdot \boldsymbol{x}^S_\theta + y^S_\theta = 0. \tag{56}$$

The planner chooses public good provision $\boldsymbol{g}_\theta \in \mathbb{R}^N_+$, restricted for simplicity to the taxable goods. This enters preferences additively separably through $\tilde{u}^h(\boldsymbol{g}_\theta, \theta)$.

Although producer prices are part of the cognition state, invariance removes their separate effect on the household problem at the consistent prior. Thus, for the purposes of Proposition 5, household demands can be represented as functions of the consumer-price system $\boldsymbol{Q} = (\boldsymbol{q}_\theta)_{\theta \in \Theta}$, with the prior

argument suppressed exactly as in Section 3.4. Write $\boldsymbol{W} \equiv (\boldsymbol{W}^h)_{h\in\mathscr{H}}$ for the given wealth, or numeraire endowment, profile, and let $\bar{W}_\theta \equiv \sum_{h\in\mathscr{H}} \mu^h w^h_\theta$ denote aggregate wealth in state θ. Let $\bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}})$ and $\bar{Y}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}})$ denote aggregate household demands at the consistent prior.[32] At the interior allocation considered below, market clearing requires

$$\bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) + \boldsymbol{g}_\theta = \boldsymbol{x}^S_\theta \qquad \text{and} \qquad \bar{Y}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) = \bar{W}_\theta + y^S_\theta.$$

The household budget constraints are already embedded in these demand functions. Using budget exhaustion to substitute out aggregate numeraire demand, numeraire-market clearing becomes

$$y^S_\theta + \boldsymbol{q}_\theta \cdot \bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) = 0, \tag{57}$$

We ignore the government budget constraint: once household budgets, market clearing, and zero profits are imposed, it is redundant by Walras' law. Indeed, (56), $\boldsymbol{x}^S_\theta = \bar{\boldsymbol{X}}_\theta + \boldsymbol{g}_\theta$, and (57) imply

$$\big(\boldsymbol{q}_\theta - \boldsymbol{p}_\theta\big) \cdot \bar{\boldsymbol{X}}_\theta = \boldsymbol{p}_\theta \cdot \boldsymbol{g}_\theta.$$

Let $\pi_\theta \lambda_\theta$ denote the multiplier on (57), let $\pi_\theta \lambda_\theta \boldsymbol{m}_\theta$ denote the multiplier on the market-clearing constraint, and let $\pi_\theta \eta_\theta$ denote the multiplier on the technology constraint. This keeps the normalization of λ_θ consistent with the main text. The planner's Lagrangian is then

$$\begin{aligned}\mathscr{L} = &\sum_{h\in\mathscr{H}} \mu^h \beta^h V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) + \sum_{\theta\in\Theta} \pi_\theta \sum_{h\in\mathscr{H}} \mu^h \beta^h \tilde{u}^h(\boldsymbol{g}_\theta,\theta) \\ &+ \sum_{\theta\in\Theta} \pi_\theta \lambda_\theta \big[y^S_\theta + \boldsymbol{q}_\theta \cdot \bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) \big] \\ &+ \sum_{\theta\in\Theta} \pi_\theta \lambda_\theta \boldsymbol{m}_\theta \cdot \big[\boldsymbol{x}^S_\theta - \bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) - \boldsymbol{g}_\theta \big] - \sum_{\theta\in\Theta} \pi_\theta \eta_\theta H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta).\end{aligned} \tag{58}$$

We first differentiate with respect to consumer prices, holding the production plan fixed. This gives

$$\sum_{h\in\mathscr{H}} \mu^h \beta^h \frac{d}{d\boldsymbol{q}_\theta} V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) + \pi_\theta \lambda_\theta \bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) + \sum_{\theta'\in\Theta} \pi_{\theta'} \lambda_{\theta'} \Big(\frac{d}{d\boldsymbol{q}_\theta} \bar{\boldsymbol{X}}_{\theta'}(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) \Big)^\top \big(\boldsymbol{q}_{\theta'} - \boldsymbol{m}_{\theta'} \big) = \boldsymbol{0}. \tag{59}$$

Once invariance is imposed, this is the same first-order condition as in Appendix B.3, except that the

[32] That is, $\bar{\boldsymbol{X}}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) \equiv \sum_{h\in\mathscr{H}} \mu^h \boldsymbol{X}^h(\theta,\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}})$ and $\bar{Y}_\theta(\boldsymbol{Q},\boldsymbol{W},\phi^*_{\boldsymbol{Q}}) \equiv \sum_{h\in\mathscr{H}} \mu^h Y^h(\theta,\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}})$.

multiplier $\boldsymbol{m}_{\theta'}$ appears in place of the producer price vector.

Under invariance, the fictitious attentive representation used in Section 3.4 applies at the consistent prior $\phi^*_{\boldsymbol{Q}}$. Thus

$$V^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) = \tilde{v}^h(\boldsymbol{Q},\boldsymbol{W}^h) \qquad \text{and} \qquad \boldsymbol{X}^h(\boldsymbol{Q},\boldsymbol{W}^h,\phi^*_{\boldsymbol{Q}}) = \big(\tilde{\boldsymbol{x}}^h(\theta,\boldsymbol{Q},\boldsymbol{W}^h)\big)_{\theta\in\Theta}.$$

Repeating the same Roy-identity and Slutsky steps as in Appendix B.3 yields

$$\sum_{h\in\mathscr{H}} \mu^h \left[\big(\pi_\theta \lambda_\theta - \widetilde{\gamma}^h_\theta\big)\bar{\boldsymbol{x}}^h_\theta + \sum_{\theta'\in\Theta} \pi_{\theta'}\lambda_{\theta'} \big(\boldsymbol{S}^h_{\theta,\theta'}\big)^\top \big(\boldsymbol{q}_{\theta'} - \boldsymbol{m}_{\theta'}\big) \right] = \boldsymbol{0}, \tag{60}$$

where $\bar{\boldsymbol{x}}^h_{\theta'} \equiv \tilde{\boldsymbol{x}}^h(\theta',\boldsymbol{Q},\boldsymbol{W}^h)$ and

$$\widetilde{\gamma}^h_\theta \equiv \beta^h \partial_{w_\theta} \tilde{v}^h + \sum_{\theta'\in\Theta} \pi_{\theta'}\lambda_{\theta'}(\boldsymbol{q}_{\theta'} - \boldsymbol{m}_{\theta'}) \cdot \partial_{w_\theta} \bar{\boldsymbol{x}}^h_{\theta'}.$$

The additional first-order conditions are with respect to the production plan and public good provision:

$$\lambda_\theta \boldsymbol{m}_\theta = \eta_\theta \nabla_{\boldsymbol{x}} H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta), \tag{61}$$

$$\lambda_\theta = \eta_\theta \nabla_y H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta), \tag{62}$$

$$\lambda_\theta \boldsymbol{m}_\theta = \sum_{h\in\mathscr{H}} \mu^h \beta^h \nabla_{\boldsymbol{g}} \tilde{u}^h(\boldsymbol{g}_\theta, \theta). \tag{63}$$

Combining (61), (62), and (55) gives

$$\boldsymbol{m}_\theta = \frac{\nabla_{\boldsymbol{x}} H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)}{\nabla_y H(\boldsymbol{x}^S_\theta, y^S_\theta, \theta)} = \boldsymbol{p}_\theta.$$

Hence the relevant wedge in (60) is

$$\boldsymbol{q}_{\theta'} - \boldsymbol{m}_{\theta'} = \boldsymbol{q}_{\theta'} - \boldsymbol{p}_{\theta'} = \boldsymbol{\tau}_{\theta'}.$$

So $\widetilde{\gamma}^h_\theta$ coincides with γ^h_θ from Section 3.4, and (60) reduces exactly to (22). If attention costs are state separable as well, it further collapses to (20). Endogenous production therefore leaves the consumer-tax formulas unchanged. It only adds the equilibrium conditions that determine producer prices and

public good provision. In particular, (63) becomes

$$\lambda_\theta \boldsymbol{p}_\theta = \sum_{h \in \mathscr{H}} \mu^h \beta^h \nabla_{\boldsymbol{g}} \tilde{u}^h(\boldsymbol{g}_\theta, \theta),$$

which is the additional condition reported in Proposition 5.

The simplified corollaries also impose the additional restriction that $\lambda_\theta = \lambda$ for all θ. With endogenous public-good provision, this is a restriction on the public-good utility primitives: the functions $\tilde{u}^h(\cdot, \theta)$ must be such that, at the optimal provision levels,

$$\sum_{h \in \mathscr{H}} \mu^h \beta^h \nabla_{\boldsymbol{g}} \tilde{u}^h(\boldsymbol{g}_\theta, \theta) = \lambda \boldsymbol{p}_\theta \qquad \forall \theta.$$

Under this restriction, and maintaining the other assumptions of Corollaries 1 and 2, the simplified formulas follow as before.

D.2 Proof of Theorem 3

[To be added.]

References

Allcott, Hunt, Benjamin Lockwood, and Dmitry Taubinsky, "Ramsey Strikes Back: Optimal Commodity Taxes and Redistribution in the Presence of Salience Effects," *AEA Papers and Proceedings*, May 2018, *108*, 88–92. 1, 1

Angeletos, George-Marios and Alessandro Pavan, "Policy with Dispersed Information," *Journal of the European Economic Association*, 2009, *7* (1), 11–60. 2

_ **and Karthik A Sastry**, "Inattentive economies," *Journal of Political Economy*, 2025, *133* (7), 000–000. 1, 1, 1, 3, 11, 14, 16, 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, 4.3

_ **,** _ **, 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, 16

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, August 2009, *99* (4), 1145–1177. 1, 1, 1, 4.3

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

Goldin, Jacob, "Optimal tax salience," *Journal of Public Economics*, November 2015, *131*, 115–123. 1, 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, 14, 16

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

Mullainathan, Sendhil, Joshua Schwartzstein, and William J Congdon, "A reduced-form approach to behavioral public finance," *Annu. Rev. Econ.*, 2012, *4* (1), 511–540. 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