Optimal Tax Salience

Jacob Goldin*

First Draft: March 2011

This Draft: November 2013

Abstract

Recent empirical work suggests that consumers systematically misperceive commodity taxes when the after-tax price is not prominent. I show how policymakers may utilize such low-salience taxes to enhance consumer welfare. The optimal combination of high- and low-salience taxes balances two competing welfare effects: low-salience taxes accommodate lower tax rates but induce consumers to misallocate their budgets. The efficiency gains from implementing the optimal policy are substantial, up to the entire deadweight loss from distortionary taxation. The surprising result that the optimal policy is to induce taxpayer mistakes can be readily understood as an application of the theory of the second-best.

*Princeton University and Yale Law School. For helpful comments, I am grateful to Constantine Angiridis, Raj Chetty, Mikhail Golosov, Jonah Gelbach, Nikolaj Harmon, Tatiana Homonoff, Louis Kaplow, Yair Listokin, David Lee, Daniel Markovits, Alex Mas, Daniel Reck, and Dean Spears, as well as participants in workshops at Princeton and Yale. All errors are my own.
1. Introduction

The subject of optimal commodity taxation lies at the heart of public finance. Most research in the area asks how policymakers should levy taxes across goods to achieve efficiency and distributional goals when lump-sum taxes are unavailable. In contrast, the question of how to implement a particular tax on a single good has not received the same degree of theoretical attention.

Recent empirical findings suggest a need to reconsider this emphasis. A series of papers suggests that the design of a tax in particular the taxes salience has important effects on consumer behavior: the more prominent the after-tax price of a good, the more consumers respond to changes in the tax rate on that good.\(^1\)

The presence of such salience effects suggests an additional margin through which policymakers can shape the behavioral effects of a tax. In many contexts, policymakers have a range of options for how to design a tax, and each option may be associated with a different degree of salience. For example, road tolls can be collected manually by cash transfers or automatically through an EZ-Pass system (Finkelstein 2009); property tax payments may be collected on their own or bundled into a monthly mortgage payment to an escrow account (Hayashi 2012, Cabral and Hoxby 2010); income tax payments may be collected from employees or automatically withheld (Jones 2010). In the context of commodity taxation, policymakers may manipulate salience through their choice of whether to include a tax in a

\(^1\)See, for example, Chetty, Looney, and Kroft (2009) (grocery store customers reduce demand for goods when the sales tax inclusive price is posted; beer consumption declines more in response to excise tax changes (high-salience) than to sales tax changes (low-salience)); Finkelstein (2009) (drivers’ behavior becomes less sensitive to tolls upon adoption of EZ-Pass systems); Cabral and Hoxby (2010) (property taxes are lower in jurisdictions that allow the tax to be collected in less salient ways); Hayashi (2012) (low-salience property tax designs reduce the probability of appealing municipal valuations); Feldman and Ruffle (2013) (lab experiment documenting participants’ tendency to spend more when faced with tax-exclusive prices as compared to tax-inclusive prices); Gallagher and Muehlegger (2008) (income tax incentives to purchase fuel-efficient vehicles were most effective in the quarters most closely following income tax payment); Fochmann and Weimann (2011) (field experiment participants exerted more effort in response to a constant net wage when taxed than when untaxed); Goldin and Homonoff (2013) (high-income consumers exhibit inattention to cigarette sales taxes). Krishna and Slemrod (2003) and McCaffery (1994) review earlier evidence on how tax design affects behavior and discuss implications for policy. A related literature documents that income tax filers tend to systematically misperceive marginal tax rates, e.g., Liebman and Zeckhauser (2004).

\(^2\)Throughout, I employ salience to refer to the prominence of the taxed good’s tax-inclusive price. Thus an excise tax included in a good’s posted price is “high-salience” even though consumers may not be able to identify how much of what they pay is tax as opposed to price charged by the retailer.
goods posted price or to add it on at the register (at the time the consumer checks out of
the store). To the extent that the former are more salient than the latter (Chetty, Looney,
and Kroft 2009), governments may alter a tax’s salience by adjusting the degree to which it
relies on these two tax designs. Thus although policymakers typically lack perfect control
over the salience of a given tax, they frequently face a choice between relying on high- and
low-salience ways of raising revenue. And because so many questions of tax design are linked
to issues of salience, policymakers end up making decisions about salience – whether they
intend to or not – whenever they levy a new tax.

This paper considers the question of optimal commodity tax salience: how should a
benevolent government choose between high- and low-salience taxes on a particular good
in order to raise some required amount of revenue? The analysis highlights two distinct
mechanisms through which tax salience affects consumers’ well-being. On the one hand, low-
salience taxes dampen the excess burden traditionally associated with distortionary taxation;
because consumers are less prone to substitute away from goods subject to low-salience taxes,
such taxes are less distortionary for a given amount of revenue raised. On the other hand,
low-salience taxes drive taxpayers to make optimization errors, reducing welfare by causing
the misallocation of income among consumption goods. The government’s decision between
high- and low-salience taxes trades off between these two welfare effects.

My results suggest an important role for salience considerations in the design of tax
policy. When policymakers lack access to lump-sum taxation, commodity taxes generate
an excess burden by distorting consumption decisions between taxed and untaxed goods.
However, I show that when the government has access to tax instruments that differ in

3Policymakers may also manipulate commodity tax salience by adopting tax-inclusive pricing regulations,
which require retailers to include the full amount of consumption taxes in the prices displayed to consumers,
for all goods. Such regulations are common in Europe but are quite rare in the United States. Similarly,
governments may require tax-inclusive pricing for a particular good. For example, new rules require websites
selling airline tickets to include state and federal airline taxes in the initial price displayed to consumers.

4To take a recent example, the state of Colorado is deciding how much of its new marijuana tax to levy
as an excise tax and how much to levy as a sales tax (New York Times 2013).

5Addressing this question through the lens of economic theory is complicated by the conceptual and
methodological challenges that arise in the context of behavioral welfare analysis. To analyze these thorny
issues, I follow Chetty et al. (2009) and adopt a refinement approach to welfare analysis, drawing on insights
from Bernheim and Rangel (2009) and Chetty (2009b). As I elaborate below, this framework allows one to
conduct welfare analysis while remaining agnostic about the exact mechanism driving taxpayer under-reaction
to low-salience taxes, at least within a broad range of plausible models.
their salience, it can employ those instruments in combination to enhance efficiency. In fact, when taxpayer mistakes are sufficiently robust to changes in the tax rate, I show that the optimal policy achieves the first-best welfare outcome, even without access to a lump-sum tax. Intuitively, the presence of multiple tax instruments with differing degrees of salience provides policymakers with an additional degree of freedom with which to shape consumer consumption decisions. The optimal combination of high- and low-salience taxes induces taxpayers to consume at the same consumption bundle they would choose if faced with a lump-sum tax (even though this consumption decision is actually sub-optimal from the consumers perspective given the distortionary tax that is actually in place). The model and assumptions I rely on to reach this result are exactly those adopted by Chetty et al. (2009).

I next turn to the question of characterizing the optimal combination of high- and low-salience taxes. Solving the government’s problem yields a simple and intuitive formula for the optimal policy. Several interesting results emerge. Most notably, the optimal size of the low-salience tax is always non-zero. Although low-salience taxes drive consumers to make optimization errors, the welfare costs of those errors is second-order for small values of the tax. In contrast, even small values of a low-salience tax may raise substantial revenues, allowing the government to reduce distortionary high-salience taxes while still meeting its budget constraint. Additionally, the optimal ratio of high- to low-salience taxes depends on the properties of the good being taxed. For luxury goods, or for goods that make up a large share of consumer expenditures, governments should (all else equal) rely more heavily on high-salience taxes. In contrast, the presence of many close substitutes for a taxed good implies that low-salience taxes will tend to be more efficient. Intuitively, high-salience taxes tend to be more efficient when consumption of the taxed good is associated with negative externalities.

Finally, I consider an extension of the model to the case in which tax salience is endogenously related to the size of the tax and derive conditions under which teh first-best welfare outcome will be attainable. Even when the first-best is unattainable, I show how policymakers may use the optimal salience formula to make incremental improvements to consumer
welfare.

Despite the ubiquity of policy decisions that affect tax salience, the topic has received little theoretical attention. Indeed, as Congdon, Kling, and Mullainathan (2009) conclude in their review of the behavioral tax literature, [T]he theoretical literature has yet to yield the type of rules of thumb with respect to optimal tax salience that translate into practical policy recommendations. Apart from the empirical work cited above, the theoretical papers closest to the current analysis are Chetty et al. (2009) and Chetty (2009a). Those authors derive formulas for measuring the excess burden of a new tax that is less than fully salient. Although closely related to the current analysis, such formulas address a distinct question and consequently yield different insights. A number of influential papers have investigated how cognitive biases other than salience affect the prescriptions for optimal tax policy (E.g., Liebman and Zeckhauser 2004, O’Donoghue and Rabin 2003). Such papers typically reevaluate the optimal level of a tax instrument conditional on taxpayers exhibiting an assumed behavioral bias. I build on this literature by considering a setting in which the government’s choice of tax instrument controls the extent to which taxpayers exhibit a bias in the first place.

The remainder of this paper proceeds as follows. Section 2 elaborates on the refinement approach to behavioral welfare analysis employed here. Section 3 develops the model and derives the basic results. In Section 4, I discuss the possibility of endogenous salience and its implications for the model’s results. Section 5 extends the results to the case in which the taxed good generates a consumption externality. Section 6 concludes.

2. Framework for Behavioral Welfare Analysis

The empirical literature on salience suggests that decision-makers fail to perfectly optimize when taxes are less than fully-salient.\(^6\) Adjusting decision-making models to account for

\(^6\)Of course, the literature discussed in Section 1 does not prove that low-salience taxes induce consumers to behave sub-optimally, even putting aside questions of econometric validity. After all, any observed choice behavior can be rationalized by sufficiently relaxing one’s assumptions concerning the content of the preferences being observed. The observed behavioral discrepancy between high- and low-salience taxes only constitutes evidence of irrationality if one assumes that the form of a tax is irrelevant to the taxpayers true preferences.
departures from rationality can have important implications for welfare analysis, but one quickly runs into the problem that a large (potentially very large) number of models may be able to explain some observed pattern of decision-making, yet each do so in a way that implies a different welfare conclusion.

One approach to addressing this problem is to attempt to differentiate competing positive behavioral models empirically, waiting to draw welfare conclusions until the true decision-making model has been identified. However, the observationally-distinct qualities of competing decision-making models can be subtle at best, and the large number of potential models available may mean that confidently identifying the true underlying model may be a Sisyphean task.\(^7\)

Instead, I pursue a “sufficient statistics” approach to the problem (discussed in Chetty 2009b). In particular, I adopt the same two assumptions relied on by Chetty et al. (2009):

(A1) Consumption is a sufficient statistic for utility: tax design affects utility only through its effect on the agents consumption.

(A2). When tax-inclusive prices are fully-salient, the agent chooses the same allocation as an optimizing agent.

Assumption (A1) states that if one holds an agents consumption bundle fixed, changing the tax rate or design does not affect the agents welfare. Although it is usually implicit, (A1) appears in most standard public finance models; writing an agent’s utility in terms of consumption alone implies that other parameters (such as the tax rate) do not directly enter the utility function. To understand the assumption, consider an example in which it fails: an agent would violate (A1) if she preferred facing a register tax to a posted tax on political grounds, perhaps because the amount being paid to the government is more transparent under the former relative to the latter. An important limitation of (A1) is that it rules out a class of interesting models in which decision-makers experience a psychic cost when faced with a low-salience tax, independent of the ultimate effect of the tax on their consumption.\(^8\)

---

See (A1) below.

\(^7\)For example, Chetty et al. (2007) develop several positive behavioral models that generate salience effects.

\(^8\)However, not all psychic cost models are ruled out. For example, suppose that accounting for a low-
Consequently, if low-salience taxes generate substantial disutility for consumers independent of their effect on consumption decisions, the results presented here will over-state the benefits of low-salience taxes by neglecting such costs. However, as Chetty et al. (2007) demonstrate, even relatively small cognitive costs generate substantial under-reaction to a tax; as a result, omitting such costs from the model may not be as misleading as one might otherwise believe.

Assumption (A2) represents a weakening of the usual instrumental rationality assumption underlying the revealed preference approach to welfare analysis. Rather than assume that all of a decision-makers choices reflect their true preferences, (A2) posits rationality only for the subset of observed choices made when taxes are fully salient. One can understand (A2) as a “refinement” within the Bernheim and Rangel (2009) framework for behavioral welfare analysis: because we have reason to be skeptical about the quality of choices made when taxes are less than fully-salient, we privilege the choices revealed when such conditions are not present.

Although this approach to behavioral welfare analysis requires imposing assumptions on the content of consumers preferences, the payoff to that assumption is substantial. I am able to derive simple and intuitive formulas for optimal tax salience by specifying only that consumers under-react to some taxes relative to others, without having to make assumptions about exactly why they under-react the way in which they do.

3. The Model and Results

Setup

Society is composed of a representative taxpayer, who divides her income between two goods: $x$ and a composite of all other goods, $y$. I assume that utility depends only on consumption of $x$ and $y$, and that the function is concave and smooth with respect to each good.$^9$ The salience tax requires a consumer to suffer some cognitive cost, but because of that cost, the consumer rationally chooses to ignore the tax. This agent does not violate (A1) because given her decision-making strategy, she does not suffer any direct utility cost when confronted with the tax. $^9$

$^9$Consistent with (A1), this implies that tax salience does not affect utility apart from its effect on consumption.
government must raise revenue $R_0$ from taxes on $x$ (good $y$ is left untaxed). The government has two tax designs from which to choose: a high-saliency tax $t_h$ and a low-saliency tax $t_l$.

The taxpayer’s budget constraint takes the form

$$y + (p + t_h + t_l) x = I$$

where $p$ represents the pre-tax price of $x$, $I$ is income, and the price of $y$ is normalized to 1. Production of $x$ is characterized by constant returns to scale technology, so that the pre-tax price of $x$ is fixed at its (constant) marginal cost. Taking income as fixed, demand for $x$ and $y$ can be expressed as a function of the two tax rates: $x = x(t_h, t_l)$ and $y = y(t_h, t_l)$. Total government revenue $R$ is thus given by

$$R(t_h, t_l) = (t_h + t_l) x(t_h, t_l)$$

Consistent with the behavioral evidence described in the introduction, the model takes as its starting point the observation that consumers adjust their demand more strongly in response to changes in high salience taxes than to changes in low salience taxes. As in Chetty et al. (2009), I adopt a functional definition of tax salience: tax salience measures how taxpayers adjust their demand for the taxed good in response to a change in the tax relative to a chance in the good’s posted price:

$$\theta^H = \frac{\partial x}{\partial t_h} \frac{\partial t_h}{\partial p}, \quad \theta^L = \frac{\partial x}{\partial t_l} \frac{\partial t_l}{\partial p}$$

To illustrate the notation, a tax that appears as part of a good’s posted price (e.g. an excise tax) would be fully-salient (i.e. $\theta = 1$). In contrast, a tax to which consumers were entirely unresponsive would imply $\theta = 0$. Equation 3 takes the place of a fully-specified positive model of how consumers perceive low salience taxes and how they adjust their demand for other goods to make their budget constraints hold. The upside to the reduced-form nature of the model is its generality; under (A1) and (A2), $\theta$ constitutes a sufficient statistic for the underlying positive model for the purposes of welfare analysis (Chetty et al. 2009). Although
I motivate and discuss the model in terms of salience, \( \theta \) may reflect any characteristic of the tax that causes individuals to systematically err in the extent to which they account for the tax.\(^{10}\)

Finally, I assume that the two taxes available to the policymaker have differing (but individually fixed) degrees of salience:

\[(A3) \ 0 \leq \theta^L < \theta^H \leq 1\]

Whether (A3) applies in a particular context is an empirical question. One common situation in which (A3) will be satisfied for commodity taxes is when the government has access to one tax instrument that is less than fully salient (\( \theta^L < 1 \)), and another that directly affects the posted price of the taxed good (\( \theta^H = 1 \)). For ease of interpretation, (A3) also imposes that the salience of the two tax instruments is between 0 and 1.\(^{11}\)

**Proposition 1**

*Under (A1) - (A3), the optimal combination of high- and low-salience taxes achieves the same welfare outcome as a non-distortionary lump-sum tax.*

**Proof of Proposition 1**

From (2), the government’s revenue constraint is given by

\[(t_h + t_l) \cdot x(t_h, t_l) = R_0\] \((4)\)

Consider a government that meets (4) by imposing some combination of high- and low-salience taxes. This government may adjust the taxes it imposes, but it must do so in a revenue-neutral way in order for (4) to continue to hold. Note that because \( \theta^H > \theta^L \), the revenue brought in by a high-salience tax increase is less than the revenue raised by a

---

\(^{10}\)As in Chetty et al. (2009), 3 implicitly assumes a constant degree of under-reaction to a less-than-fully salient tax. As discussed below, additional feasibility considerations arise if the salience of a tax is endogenously related to other variables in the model.

\(^{11}\)This restriction is also substantive in that some of the following results do not hold when the salience of both tax instruments exceed 1.
low-salience tax increase of the same magnitude: \(^{12}\)

\[
\frac{\partial R}{\partial t_l} = \theta^L \frac{\partial x}{\partial p} (t_h + t_l) + x > \theta^H \frac{\partial x}{\partial p} (t_h + t_l) + x = \frac{\partial R}{\partial t_h}
\]

Consequently, a revenue-neutral increase in \(t_l\) accommodates a greater than one-for-one reduction in \(t_h\):

\[
\left. \frac{\partial t_h}{\partial t_l} \right|_{R_0} = -\frac{\theta^L \frac{\partial x}{\partial p} (t_h + t_l) + x}{\theta^H \frac{\partial x}{\partial p} (t_h + t_l) + x} < -1 \tag{5}
\]

Now consider the welfare effect of a revenue-neutral increase in \(t_l\) and reduction in \(t_h\). The agent’s utility induced by a particular tax policy is given by

\[
V(t_h, t_l) = U(x(t_h, t_l), y(t_h, t_l)) \tag{6}
\]

Totally differentiating this function reveals that a revenue-neutral shift towards the low salience tax will benefit consumers if and only if

\[
\left. \frac{dV}{dt_l} \right|_{R_0} \geq 0 \iff U_x(x, y) \left|_{R_0} + U_y(x, y) \left|_{R_0} + U_y(x, y) \right|_{R_0} \geq 0 \tag{7}
\]

Drawing on our behavioral assumptions (3) and the consumer’s budget constraint (1), we can rewrite (7) as:

\[
\left. \frac{dV}{dt_l} \right|_{R_0} \geq 0 \iff -x \left( 1 + \left. \frac{\partial t_h}{\partial t_l} \right|_{R_0} \right) U_y(x, y) + \left. \frac{\partial x}{\partial p} \left( \theta^L + \left. \frac{\partial t_h}{\partial t_l} \right|_{R_0} \right) \theta^H \right) (U_x(x, y) - (p + t_h + t_l) U_y(x, y)) \geq 0 \tag{8}
\]

In (8), the first term represents the positive utility gain stemming from the additional purchasing power associated with the lower (total) taxes made possible by the shift. The second term captures the utility loss from the additional mistake induced by the shift, i.e. the

\(^{12}\)I assume throughout that demand for the taxed good is not so sensitive that levying a new excise tax would actually reduce revenue: \(\frac{\partial x}{\partial p} (t_h + t_l) + x > 0.\)
departure from the optimal allocation between $x$ and $y$.

Finally, using (5) and a little algebra, it is straightforward to show that (8) simplifies to:

$$\left. \frac{dV}{dt} \right|_{R_0} \geq 0 \iff U_x(x, y) - p U_y(x, y) \geq 0 \quad (9)$$

Note that at the optimum, the government’s choice of $t_h$ and $t_l$ satisfies (9) and (4) with equality. Note further that the well-behaved nature of the the utility function and revenue constraint guarantees a unique solution to the government’s maximization problem. That is, Equations (1), (9), and (4) uniquely determine the optimal tax combination by pinning down the taxpayer’s consumption of $x$ and $y$.

To prove Proposition 1, we will compare the conditions characterizing the solution to the government’s optimal salience problem with the conditions characterizing the taxpayer’s consumption under a lump-sum tax. Consider the effect on consumption of a (fully-salient) lump-sum tax of size $R_0$. Because the lump-sum tax is fully-salient (A2) implies that we can find the consumption bundle induced by the tax by solving the ordinary consumer maximization problem subject to the lump-sum tax budget constraint:

$$p x_{LST} + y_{LST} = I - R_0 \quad (10)$$

which yields the first-order condition:

$$U_x(x_{LST}, y_{LST}) - p U_y(x_{LST}, y_{LST}) = 0 \quad (11)$$

Comparing the conditions that characterize consumption under the optimal salience policy $- 1$, (4) and (9) – with the conditions that characterize consumption under the lump-sum tax ((10) (11)) reveals that both tax instruments induce the same consumption bundle. Because (A1) guarantees that consumption is a sufficient statistic for welfare, the optimal salience policy and the lump-sum tax achieve the same welfare result as well.
Discussion of Proposition 1

Proposition 1 states that when policymakers can manipulate the design of distortionary commodity taxes to modulate consumer responsiveness, they can do so in a way that replicates the consumption decisions that taxpayers would make under a lump-sum tax and that raises the same amount of revenue.

To understand the intuition behind the result, it is helpful to consider a stylized example. Suppose that the government is choosing between a fully-salient tax on $x$ $(\theta^H = 1)$ and a tax to which consumers are entirely unresponsive $(\theta^L = 0)$. The consumer’s pre-tax budget constraint is given by AB in Figure 1, and consumption $(x_0)$ is characterized by the tangency of the consumer’s indifference curve $(IC_0)$ with AB. Because any feasible choice of tax rates must raise $R_0$, the taxpayer’s final consumption will lie somewhere on the line CD, which is simply AB shifted down by the vertical distance $R_0$.

If the government relied solely on $t_h$, the consumer’s budget constraint will shift to AE; consumption of the taxed good $(x_h)$ is the value that induces tangency with the consumer’s new indifference curve $(IC_h)$. As in the standard case, the tax generates excess burden by driving consumers to substitute away from the taxed good.

In contrast, if the government relied solely on $t_l$, consumption of the taxed good $(x_l)$ would not change: $x_l = x_0$. Although consumers do not substitute away from the taxed good, the tax still generates an excess burden because consumers fail to adjust their consumption to account for the tax’s (non-distortionary) income effect.

Finally, under a lump-sum tax, there are no relative price changes. Consequently, consumption of the taxed good $(x_{LST})$ only falls in response to the tax’s income effect. Graphically, $x_{LST}$ is characterized by the point at which the consumer’s indifference curve is tangent to the new budget constraint (CD). Assuming that $x$ is a normal good, it is easy to see that $x_h < x_{LST} < x_l$.\(^{13}\) Because the lump-sum tax represents the first-best welfare outcome, the optimal policy lies somewhere between full reliance on either $t_h$ or $t_l$. Intuitively, by shifting

\(^{13}\)This condition may fail when $\theta^H < 1$ or $\theta^L > 0$. In such cases, the government may need to rely on a combination of taxes and subsidies to reach $x_{LST}$. If subsidies are unavailable, the optimal policy takes the form of a corner-solution, discussed below.
the balance between the high- and low-salience tax, the government can move consumption along CD until it achieves $x_{LST}$.

**Corollary 1.1**

Suppose that $\theta^H = 1$. Then under $(A1) - (A3)$, the optimal size of the low-salience tax is non-zero, $t_l \neq 0$.

**Proof of Corollary 1.1**

By contradiction. Suppose that the optimal combination of high- and low-salience taxes entailed $t_l = 0$. Under (A2), this policy implies that the taxpayer behaves as a fully-optimizing agent. Because the taxpayer’s budget constraint is given by $(p + t_h) x+y=I$, consumption satisfies the standard first-order condition for an interior maximum:

$$U_x(x, y) - U_y(x, y) (p + t_h) = 0$$  \hspace{1cm} (12)

However, we also know that at the optimal combination of high- and low-salience taxes, consumption is such that 9 holds with equality. Because $R_0 > 0$ implies $t_h > 0$, Equations 9 and 12 may not be satisfied simultaneously. Hence we may conclude that $t_l \neq 0$.

**Discussion of Corollary 1.1**

Corollary 1.1 highlights a somewhat surprising result. Even when policymakers have access to a fully-salient tax instrument – one that induces no consumer mistakes – consumers are actually better off when the government utilizes a less than fully-salient tax. Intuitively, the optimal tax salience problem reflects a basic tension between high- and low-salience taxes. On the one hand, the less salient a tax is, the more it mutes consumer substitution away from the taxed good, thereby reducing the deadweight loss typically associated with non-lump-sum taxes. On the other hand, by causing consumers to depart from optimal decision-making, low salience taxes drive consumers to make optimization errors when making their purchasing decisions.
The key insight is that when taxes on $x$ are close to fully-salient, the former effect will be large relative to the latter. To see why, recall that by making the tax a little less salient, the government can raise the same amount of revenue while reducing the tax’s distortionary effects on consumption, thereby reducing the traditional source of excess burden. Although the reduction in salience does drive consumers to accidentally over-consume $x$ relative to $y$, the utility cost of that optimization error is trivial when the tax is close to fully salient; because consumers facing a fully-salient tax align the marginal utility of expenditures on $x$ and $y$, consuming a little too much $x$ relative to the optimum will not generate much less utility than if the consumer had purchased $y$ instead. Put differently, because the optimization error engendered by the low-salience tax depends on the difference in marginal utilities between $x$ and $y$, the welfare cost of that error is small when the consumer is near the optimum, (i.e. when the tax is close to fully salient).

More practically, Corollary 1.1 highlights that whenever taxes on a particular good are fully-salient, welfare can be improved by a small revenue-neutral shift towards a less-salient tax instrument.

**Proposition 2: Implementing the Optimal Solution**

Define $\eta_{x,I} = \frac{\partial r}{\partial I} (x)$ (income-elasticity of $x$), $\omega_x = \frac{p_x}{T}$ (budget share of expenditures on $x$), and $\varepsilon_{x,p} = -\frac{\partial x}{\partial p} (x)$ (own-price elasticity of $x$, defined to be positive).

Let $\rho$ denote the fraction of taxes on $x$ that are low-salient: $\rho \equiv \frac{t_i}{t_h+t_i}$. Under (A1) - (A3), the optimal combination of high- and low-salience taxes is (approximately) given by the value of $\rho$ that solves $\rho \theta^L + (1 - \rho) \theta^H = \theta^*$, where $\theta^* = \frac{\eta_{x,I} \omega_x}{\varepsilon_{x,p}}. \quad 14$

Additionally, define $\tau$ to be the size of the commodity tax needed to generate $R_0$ when taxpayers consume at the first-best level of $x$, $\tau = \frac{R_0}{x_{LST}}$. Define $t_h$ to be the level of the high-salience tax that induces taxpayers to consume at the first-best level of $x$ (when no other taxes are imposed): $x(t_h,0,1) \equiv x_{LST} \equiv x(0,0,1 - R)$. The values of the high- and low-salience taxes that achieves the first best welfare outcome is (exactly) given by:

\[14\text{All quantities are evaluated at the no-tax baseline.}\]
\[ t^*_h = \frac{\theta^H \bar{t}_h - \theta^L \tau}{\theta^H - \theta^L}, \quad t^*_l = \frac{\theta^L (\tau - \bar{t}_h)}{\theta^H - \theta^L} \]

Proposition 2 states that for a given taxed good, there exists some target level of salience, \( \theta^* \), the value of which depends upon the nature of demand for the good in question. When the government can choose from tax instruments with varying salience, it can maximize efficiency by choosing among them by reference to \( \theta^* \).

**Proof of Proposition 2**

To understand how one can employ high- and low-salience taxes to implement the first-best solution, consider the following approach. First, set \( t_h = \bar{t}_h \), where \( \bar{t}_h \) is defined as the level of the high-salience tax that induces consumption of \( x \) at the first-best amount: \( x(\bar{t}_h, 0, I) \equiv x(0, 0, I - \bar{R}) \equiv x_{LST} \). Although \( \bar{t}_h \) induces the first-best level of \( x \), the revenue raised at \( \bar{t}_h \) is less than \( R_0 \) (because \( t_h \) is distortionary).

The key to meeting the government’s revenue target without departing from \( x_{LST} \) is to adjust the balance between \( t_h \) and \( t_l \) in ways that increase revenue but do not affect consumption of \( x \). That is, by strategically combining increases in \( t_l \) with reductions in \( t_h \), policymakers can increase revenue without causing individuals to substitute away from \( x \). To see this, note that totally differentiating \( x(t_h, t_l) \) yields:

\[ dx = \frac{\partial x}{\partial t_l} dt_l + \frac{\partial x}{\partial t_h} dt_h = \theta^L \frac{\partial x}{\partial p} dt_l + \theta^H \frac{\partial x}{\partial p} dt_h \]

Consequently, movements along the line \( \frac{\partial x}{\partial t_h} |_{x_{LST}} = \frac{\theta^H}{\theta^L} \) do not affect the taxpayer’s demand for \( x \). However, movements along this line can affect the amount of revenue raised. In particular, the revenue effects of reducing \( t_h \) while increasing \( t_l \) in an amount that does not affect demand for \( x \) is given by:

\[ \frac{dR}{dt_h} |_{x_{LST}} = - \left[ (t_l + t_h) \frac{\partial x}{\partial t_l} |_{x_{LST}} + x_{LST} \left( 1 + \frac{\partial t_l}{\partial t_h} |_{x_{LST}} \right) \right] = x_{LST} \left( \frac{\theta^H - \theta^L}{\theta^L} \right) \quad (13) \]
Thus for each $\$1$ reduction in $t_h$, $t_l$ may be increased by $\frac{\partial H}{\partial t}$ dollars without causing consumption of $x$ to depart from $x_{LST}$. At the same time, this policy change raises revenue in the amount of $x_{LST} \left( \frac{\partial H - \partial L}{\partial t} \right)$ dollars.

Suppose the government initially sets $(t_h, t_l) = (\overline{t}_h, 0)$ and subsequently reduces $t_h$ by $\delta$ dollars while increasing $t_l$ by $\frac{\partial H}{\partial t} \delta$ dollars. The net result of this policy is that consumers choose $x = x_{LST}$, and the total amount of revenue raised is $R = \overline{t}_h x_{LST} + \delta x_{LST} \left( \frac{\partial H - \partial L}{\partial t} \right)$. Setting $R = R_0$ allows us to solve for the value of $\delta$ that raises the required amount of revenue:

$$\delta = (\tau - \overline{t}_h) \left( \frac{\theta^L}{\theta^H - \theta^L} \right)$$

(14)

where

Using 14, we can solve for the optimal values of $t_h$ and $t_l$:

$$t_h^* = \overline{t}_h - \delta = \left( \frac{\theta^H \overline{t}_h - \theta^L \tau}{\theta^H - \theta^L} \right)$$

(15)

$$t_l^* = \frac{\theta^H}{\theta^L} \delta = (\tau - \overline{t}_h) \left( \frac{\theta^L}{\theta^H - \theta^L} \right)$$

(16)

Equations 15 and 16 allow one to implement the first-best solution exactly, given knowledge of $x_{LST}$ and $\overline{t}_h$. To implement the optimum when $\overline{t}_h$ is not known, and to better understand the mechanisms at work, we can approximate $\overline{t}_h$ as a function of more familiar quantities. By definition of $\overline{t}_h$, we have that $x(0, \overline{t}_h, I) \equiv x(0, 0, I - R_0)$. Subtracting $x(0, 0, I)$ from both sides and applying a first-order Taylor approximation yields: $\overline{t}_h \theta^H \frac{\partial x}{\partial \overline{t}} \approx -R_0 \frac{\partial x}{\partial I}$. It is then straightforward to show:

$$\overline{t}_h \approx \left( \frac{\eta_{x,I} \omega_x}{\varepsilon_{x,p}} \right) \frac{\tau}{\theta^H}$$

(17)

where $\eta_{x,I} = \frac{\partial x}{\partial I} \frac{1}{x}$ denotes the income-elasticity of $x$, $\omega_x = \frac{p_x}{T}$ denotes the budget share of expenditures on $x$, and $\varepsilon_{x,p} = -\frac{\partial x}{\partial p} \frac{1}{x}$ denotes the own-price elasticity of $x$ (defined to be positive), and all quantities are evaluated at the no-tax baseline.
Substituting 17 into 15 and 16 yields a more intuitive characterization of the optimal tax rates:

\[ t_h^* \approx \frac{\tau}{\theta^H - \theta^L} (\theta^* - \theta^L) \]  

(18)

\[ t_i^* \approx \frac{\tau}{\theta^H - \theta^L} (\theta^H - \theta^*) \]  

(19)

where \( \theta^* = \frac{\eta_{x,I} \omega_x}{\varepsilon_{x,p}} \).

Finally, Equations 18 and 19 still require knowledge of \( \tau = R_0/x_{LST} \) to implement. It is useful to derive a general rule of thumb when such quantities are not known precisely. Define \( \rho \) to be the optimal ratio of the low-salience tax to total taxes on \( x \), \( \rho = \frac{t_i^*}{t_h^* + t_i^*} \). Noting that \( t_h^* + t_i^* = \tau \), we can rewrite 19 in terms of \( \rho \):

\[ \rho \theta^L + (1 - \rho) \theta^H = \theta^* \]  

(20)

**Numerical Illustration of Proposition 2** For a concrete illustration, consider Proposition 2 in the context of a specific utility function and positive model of behavior. Suppose that the taxpayer has Cobb-Douglas utility \( u(x, y) = x^{1/3} y^{2/3} \) and sets consumption of \( x \) before choosing consumption of \( y \). The government has two tax instruments, with \( \theta^H = 1 \) and \( \theta^L = 1/6 \). Finally, assume that the pre-tax price (\( p \)) of \( x \) is 5, the agent has income (\( I \)) of 120, and the government has a revenue target \( R_0 = 30 \).

Under a lump-sum tax of 30, Marshallian demand for \( x \) is given by \( x_{LST} = x(p, I - R_0) = \frac{1}{3} \frac{I - R_0}{p} = \frac{1}{3} \frac{190}{5} - 6 \). Similarly, Marshallian demand for \( y \) is given by \( y_{LST} = y(p, I - R_0 = I - R_0 - p x = 120 - 30 - (5 \cdot 6) = 60 \).

For Proposition 1 to hold, there must exist some \((t_h, t_i)\) combination that raises revenue of 30 and induces the same consumption as the lump-sum tax, namely \((x, y) = (6, 60)\). First, note that \( \tau = R_0/x_{LST} = 30/6 = 5 \). Additionally, from the agent’s utility function, it is straightforward to solve for \( \bar{t}_h \), the value of \( t_h \) that induces \( x_{LST} = 6 \). In particular, \( \bar{t}_h = 5/3 \). Substituting these values into 15 and 16 yields \( t_h^* = 1 \) and \( t_i^* = 4 \).\(^{15}\)

\(^{15}\)To verify that these tax rates induce the first-best consumption bundle, note that given our assumptions,
Discussion of Proposition 2

Equations 18, 19 and 20 yield a number of important insights. First, \( \theta^* \) represents the optimal degree of salience for taxes on \( x \), in the following sense. When the government has available to it a tax instrument (either \( t_h \) or \( t_l \)) with salience \( \theta^* \), the optimal policy is to rely on it entirely. When no such tax is available, the government may replicate its welfare effects by employing a combination of the tax instruments that are available. The closer the salience of an available tax is to \( \theta^* \), the more heavily the government should rely on it.

Second, the \( \theta^* \) formula provides insight into how the optimal combination of high- and low-salience taxes achieves the first best; loosely speaking, it does so by replicating the behavioral effects of a lump-sum tax. That is, an increase in a fully-salient commodity tax reduces demand for the taxed good through both a substitution and an income effect, \( \varepsilon_{x,p} = \varepsilon_{x,p} + \omega_x \eta_{x,I} \), where \( \varepsilon_{x,p} \) denotes the compensated (Hicksian) own-price elasticity of demand. In contrast, an increase in a lump-sum tax reduces demand for the tax good solely through an income effect, \( \omega_x \eta_{x,I} \). By scaling the demand response to the tax by \( \theta^* = \frac{\omega_x \eta_{x,I}}{\varepsilon_{x,p}} \), the optimal policy effectively removes all but the income effect associated with the low-salience tax.\(^{16}\) Note that for normal goods, \( \theta^* \in [0,1) \).

Third, the optimal combination of high- and low-salience instruments depends upon the nature of demand for the good being taxed. In particular, \( \theta^* \) is smaller for goods with close substitutes\(^{17}\) and larger for luxury goods and for goods that constitute a large share of the taxpayer’s budget.\(^{18}\) Additionally, \( \theta^* = 0 \) if and only if demand for the taxed good is entirely insensitive to income (\( \eta_{x,I} = 0 \)).\(^{19}\) Note that although the quantities that determine \( \theta^* \) are

\[ x(p,t_h,t_l,I) = \frac{t_l/3}{p+t_h+\frac{t_l}{2}} \quad \text{and} \quad y(p,t_h,t_l,I) = I - (p + t_h + t_l)x(p,t_h,t_l,I) = I \left( \frac{2p+2t_h-t_l/2}{3p+3t_h+t_l} \right). \]

Substituting \( t_l = 4 \) and \( t_h = 1 \) into the demand functions (along with \( I = 120 \) and \( p = 5 \)) yield \( x = 6 \) and \( y = 60 \). Finally, revenue \( R = (t_h + t_l)x = (5)(6) = 30 \).

Additionally, this result is consistent with the formula derived by CLK for the deadweight loss associated with the introduction of a tax into a previously untaxed market. In particular, that formula implies that a new tax would generate no deadweight loss if it happens to have salience equal to \( \theta^* \).

\(^{16}\)Intuitively, this result follows from the fact that the excess burden associated with a tax depends on the compensated elasticity of the taxed good (Auerbach 1985). The greater the compensated elasticity, the larger the welfare gains from reducing the consumer substitution that is typically associated with commodity taxes.

\(^{17}\)These results follow from applying the Slutsky equation to write the \( \theta^* \) formula as \( \theta^* = \frac{\omega_x \eta_{x,I}}{\varepsilon_{x,p} + \omega_x \eta_{x,I}} \).

\(^{18}\)Mechanically, this result follows from the fact that \( \varepsilon_{x,p} > 0 \) when consumers behave optimally, and (A1)
readily estimable, care must be taken to evaluate them at the no-tax baseline.

Finally, Proposition 2 highlights the conditions under which subsidies will be required to implement the first-best welfare outcome. In particular, when subsidies are unavailable, the government will be able to implement the first-best welfare outcome if and only if \( \theta^* \in [\theta^L, \theta^H] \).

In that case, it is straightforward to show that the optimal policy takes the form of a corner solution, in which the government relies solely on the tax that has salience closest to \( \theta^* \).

When subsidies are available and \( \theta^* < \theta^L \), Equations 18 and 19 show that implementing the first-best requires utilizing a high-salience subsidy in conjunction with a low-salience tax. Additionally, 18 highlights the factors that shape how large the high-salience subsidy must be to achieve the first-best, for a given value of \( \theta^* \). First, when revenue requirements from taxes on \( x \) (\( R_0 \)) are large and the amount of \( x \) consumed under the first-best policy (\( x_{LST} \)) is small, \( \tau \) will be large and hence the required subsidy will tend to be large as well. Second, when the available tax instruments have similar salience, i.e. \( \theta^H \approx \theta^L \), the required subsidy will be quite large. In the extreme case in which \( \theta^H = \theta^L \), the required subsidy would be infinitely large and (A3) will not hold.

Although Proposition 1 shows that the first-best welfare outcome can technically be achieved as long as (A1) - (A3) hold, there are reasons to suspect that the model may guarantees that consumers behave optimally at the no-tax baseline where \( \tilde{z}_{x,p} \) is evaluated. To understand the intuition, consider a tax to which consumers are entirely unresponsive (\( \theta = 0 \)). Let \((x_0, y_0)\) represent the taxpayer’s initial consumption of \( x \) and \( y \) at tax \( \theta^0. \) Suppose the government raises the tax to \( t^1 = t^0 + \alpha. \) Because \( \theta = 0 \), consumers buy the same amount of \( x \) as before the tax increase, leaving them with \( \alpha x_0 \) less income to spend on other goods.

When \( \eta_{x,t} = 0 \), the consumer’s response to the tax exactly matches what a fully-optimizing agent would do. Because the optimal choice of \( x \) does not depend on income, the consumer has nothing to gain by reconsidering her consumption on \( x \) after a decline in income. In contrast, when \( \eta_{x,t} > 0 \), the consumer who fails to adjust her consumption of \( x \) in response to a tax increase is worse off for failing to do so. See Chetty et al. (2009) for a closely related discussion.

Mechanically, this follows from 18 and 19. An immediate implication is that the first-best is always attainable without subsidies when the government has available to it taxes with salience \( \theta^L = 0 \) and \( \theta^H = 1. \)

When considering subsidies, I focus on the case in which \( \theta^* < \theta^L \) rather than the case in which \( \theta^* > \theta^H \) for two reasons. First, in many contexts, the government will have a tax instrument such as an excise tax that can shape the posted price, implying \( \theta^H = 1 \geq \theta^* \). Second, numerical computations suggest that for many goods, \( \theta^* \) is likely to be quite low.

A final point relating to subsidies is that the analysis thus far has implicitly assumed that the salience of a tax instrument is identical to the salience of a similarly-designed subsidy. Of course, this need not be the case. However, such asymmetries are unlikely to pose serious problems to implementation when the government can impose a subsidy that directly affects a good’s posted price. And in at least certain contexts, tax subsidies may have higher salience than similarly-designed taxes (Feldman and Ruffle 2013).
break down when large taxes (and subsidies) are required. In particular, the model assumes that the salience of the available tax instruments is fixed and constant. In practice, there are good reasons to think that consumers will become more attentive to a tax as the amount at stake increases. Thus the feasibility of the first-best welfare outcome may be most plausible when the required size of the low-salience tax is not too large. From 19, we know this tends to occur when the difference in the salience of the available tax instruments ($\theta^H - \theta^L$) is relatively large, the revenue required to be raised from $x$ ($R_0$) relatively small, the amount of $x$ consumed under the first-best ($x_{LST}$) relatively large, and the optimal degree of salience ($\theta^*$) relatively high. The next section formalizes these intuitions.

4. Optimal Policy When Salience is Endogenous

Thus far, I have assumed that the degree of salience associated with the available tax instruments is fixed and exogenous to the model. In practice, however, it is possible that the salience of a tax depends in part on the size of the tax. For example, a bounded rationality model of decision-making suggests that consumers will be more likely to pay attention to larger taxes because the utility costs of neglecting the tax will tend to be larger as well (Chetty, Looney, and Kroft 2007). When taxpayers behave in this way, the salience of the tax will be increasing in the tax’s size.

When the salience of a tax is a function of the tax’s size, attaining the first-best may no longer be feasible, depending on the $\theta(\cdot)$ functions at hand. To see this, suppose that the

\footnote{See Reck (2013) for a formal demonstration of this claim in the context of a bounded rationality model of tax salience.}

\footnote{On the other hand, researchers have documented behavioral biases even in decision-making contexts where the stakes are large, for example retirement savings decisions (Beshears et al. 2009), high-interest borrowing (Bertrand and Morse), organ donations (Johnson and Goldstein 2003), labor supply decisions by earned income tax credit filers (Chetty and Saez 2009), and property tax assessment appeals (Hayashi 2013). In the commodity tax context, Fehrman and Ruffle (2013) do not find any evidence that tax salience effects diminish when moving from low- to high-priced products, even though a consumer’s failure to consider an ad-valorem tax has a larger utility cost in the latter case than in the former.}

\footnote{Additionally, the salience of the tax (as it is defined here) will reflect both the level of taxpayers’ attentiveness to the tax as well as the change in that attentiveness associated with the tax increase. That is, for tax t with salience $\theta$, the effect of increasing the tax is given by: $\frac{dx}{dt} = \frac{\partial x}{\partial t} + \frac{\partial x}{\partial \theta} \frac{\partial \theta}{\partial t}$. Consequently, some values of a low-salience tax may be associated with a $\theta > 1$, even when some taxpayers remain inattentive to the tax. Put differently, an increase in the degree to which consumers account for a tax (following an increase in that tax) are captured in the value of $\theta$.}
government has two tax instruments available to it, a high-salience tax with $\theta^H$ fixed at 1, and a low-salience tax where the salience depends upon the size of the tax, $\theta^L = \theta^L(t_l)$.

We can begin as before by setting $t_h$ at the level necessary to induce consumers to consume $x$ at the first-best amount, $t_h = \overline{t}_h$. Like before, consider a reduction in $t_h$ along with an increase in $t_l$ so that the net effect is to leave consumption of $x$ at $x_{LST}$. Totally differentiating demand for $x$ yields $\frac{\partial \overline{t}_h}{\partial t_l} |_{x_{LST}} = -\theta^L(t_l)$. As a result, the additional revenue generated by an “$x$-neutral” increase in $t_l$ is given by $\frac{\partial P}{\partial t_l} |_{x_{LST}} = \frac{\partial (t_l + t_h)}{\partial t_l} |_{x_{LST}} x_{LST} = \left(1 - \theta^L(t_l)\right) x_{LST}$. In order to attain the first-best, the government must be able to increase $t_l$ (and reduce $t_h$) by a sufficient amount to raise $R_0$ without altering demand for $x$. Consequently, the first-best welfare outcome is feasible if and only if there exists a value of the low-salience tax, $\hat{t}_l$, such that

$$\overline{t}_h x_{LST} + \int_0^{\hat{t}_l} \frac{\partial P}{\partial t_l} |_{x_{LST}} \, dt_l \geq R_0 \iff \overline{t}_h + \int_0^{\hat{t}_l} \left(1 - \theta^L(t_l)\right) \, dt_l \geq \tau \quad (21)$$

Thus, when the salience of the available tax instruments is endogenous, determining whether the first-best welfare outcome is feasible depends on the specifics of the $\theta^L(.)$ function over the range of values necessary to generate $R_0$. The smaller the revenue that must be raised from $x$, the larger the agent’s consumption of $x$, and the slower that $\theta^L$ increases, the more likely it is that the government will be able to attain the first-best welfare outcome.

Additionally, even when the salience of the available tax instruments increases too fast to achieve the first-best, the above results still provide local guidance to policymakers. Under (A1)-(A3), and the additional assumption of additively-separable utility, Appendix A shows that policymakers should incrementally increase their reliance on the low-salience tax whenever $\rho \theta^L + (1 - \rho) \theta^H > \theta^*$, where $\rho$ and $\theta^*$ are defined as in Proposition 2.26 Thus as long as policymakers are able to identify $\theta^L$ and $\theta^H$ at the current tax levels, they can evaluate whether a small increase in their reliance on one of the two taxes will generate efficiency benefits.27

26By increase their reliance on the low-salience tax, I mean increasing $t_l$ by a small amount while also reducing $t_h$ by the amount that leaves total revenue unchanged.

27One possibility, inconsistent with (A2), is that taxpayers suffer psychological costs from accounting for low-salience taxes, and that these costs are increasing as attentiveness to the tax increases. In this case, 21 being satisfied no longer guarantees that policymakers can reach the first-best; even when the taxes induce consumers to choose the first-best bundle of goods, consumers may be worse-off relative to the first-
5. Optimal Salience in the Presence of Externalities

Proposition 3

Suppose that consumption of $x$ generates an externality $\phi(x)$. Then under (A1) - (A3):

(a) The optimal combination of high- and low-salience taxes achieves the first-best welfare outcome.

(b) The optimal combination of high- and low-salience taxes is (approximately) characterized by $\rho \theta^L + (1 - \rho)\theta^H = \theta^{**}$, where $\rho = \frac{t_p}{t_h + t_l}$, $\theta^{**} = \theta^* + (1 - \theta^*) \frac{t_p}{\tau}$, and $t_p$ is the size of the pigouvian tax that induces the first-best when a lump-sum tax is available, $t_p = \frac{\phi'(x^*)}{U_y(x^*, y^*)}$.

Part (b) states that when consumption of $x$ generates a negative (positive) externality, the optimal degree of reliance on the high-salience tax ($t_h$) is higher (lower) than when no externality is present.

Proof of Proposition 3

For concreteness, assume that consuming $x$ reduces social welfare $W$ by some positive quantity $\phi(x)$, $\phi' > 0$, $\phi'' > 0$. The proof is analogous for the case in which $x$ generates a positive externality.

Part (a) Social welfare $W$ is given by the sum of the agent’s individual utility and the externality, $W = U(x, y) - \phi(x)$. By substituting the consumer’s demand function, we can express $W$ as a function of the tax rates:

$$W(t_h, t_l) = U(x(t_h, t_l), y(t_h, t_l)) - \phi(x(t_h, t_l))$$

(22)

The effect on social welfare of a revenue-neutral shift towards low-salience taxes is given by

---

best because they are suffering the psychological costs associated with paying some attention to the low-salience tax. On the other hand, to the extent that attentiveness to the taxes is (locally) fairly constant, policymakers may still employ the local approach described in this section for making incremental changes to the combination of high- and low-salience taxes that they employ.
\[
\frac{dW}{dt}\bigg|_{R_0} = U_x(x, y) \frac{\partial x}{\partial t} + U_x(x, y) \frac{\partial t}{\partial t} + U_y(x, y) \frac{\partial y}{\partial t} + U_y(x, y) \frac{\partial t}{\partial t} - \phi'(x) \left( \frac{\partial x}{\partial t} + \frac{\partial x}{\partial t} \frac{\partial t}{\partial t} \right)
\]

Applying that same approach as in the proof of Proposition 1, it is straightforward to show

\[
\frac{dW}{dt}\bigg|_{R_0} \geq 0 \iff U_x(x, y) - p U_y(x, y) - \phi'(x) \geq 0
\]

(23)

Of course, 23 is exactly the condition that characterizes the first-best welfare outcome in the presence of externalities: the marginal individual benefit of buying an additional unit of the taxed good, \( U_x(x^*, y^*) \), is balanced against the marginal individual and social costs associated with doing so, \( p U_y(x^*, y^*) + \phi'(x^*) \).

**Part (b)** The approach to characterizing the optimal combination of high- and low-salience taxes here is similar to the case in which externalities are not present (Proposition 2). As before, define \( t_h \) to be the value of the high-salience tax that induces the agent to consume at the first-best level of consumption, \( x_{1st} \). And as before, (A3) implies that the government can increase \( t_l \) and decrease \( t_h \) in a combination that increases revenue but leaves demand for \( x \) unchanged. Thus as before, the optimal values of the high- and low-salience taxes, \( t_h^* \) and \( t_l^* \), are given by 15 and 16. The only difference caused by the presence of the externality is in the level of \( x \) associated with the first-best, and hence in the value of \( t_h \) as well.

To find \( t_h \), first note that because of the form of the problem, the first-best solution is achieved by setting a fully-salient commodity tax equal to the marginal social cost of the externality, and employing a lump-sum tax to generate (or refund) any additional revenue that is required. That is, if we were to express demand for \( x \) as a function of the posted price, the high- and low-salience taxes, and income, we can write \( x_{1st} \equiv x(p + t_p, 0, 0, I - L) \), where \( t_p = \phi'(x_{1st})/U_y(x_{1st}, y_{1st}) \) is the pigouvian tax necessary to achieve the first-best when a lump-sum tax is available, and \( L = R_0 - t_p x_{1st} \) is the value of that lump-sum tax. The
value of $\overline{t_h}$ is thus implicitly defined by the following equation:

$$x(p + t_p, 0, 0, I - L) \equiv x(p, \overline{t_h}, 0, I)$$  (24)

To approximate the solution to the optimal combination, note that subtracting $x(p, 0, 0, I)$ from both sides of 24 and taking first-order Taylor approximations yields:

$$\overline{t_h} \frac{\partial x}{\partial p} \theta^H = t_p \frac{\partial x}{\partial I} L \frac{\partial x}{\partial I}$$

where all derivatives are evaluated at the no-tax baseline. Using the definition of $L$ and the Slutsky equation, this implies

$$\overline{t_h} = \frac{1}{\theta H} \left[ \tau \frac{\omega_x \eta_x, I}{\varepsilon_{x,p}} + t_p \frac{\xi_{x,p}}{\varepsilon_{x,p}} \right] = \frac{1}{\theta H} \left[ \theta^* \tau + (1 - \theta^*) t_p \right]$$  (25)

where the last equality follows from applying the definition of $\theta^*$ (from Proposition 2), $\theta^* = \frac{\eta_x \tau \omega_x}{\varepsilon_{x,p}}$.

Substituting 25 into 14, 15, and 16 yields formula for the optimal high- and low-salience taxes in the presence of an externality:

$$t_h^* \approx \frac{1}{\theta H - \theta L} \left[ (\theta^* - \theta^L) \tau + (1 - \theta^*) t_p \right]$$  (26)

$$t_l^* \approx \frac{1}{\theta H - \theta L} \left[ (\theta^H - \theta^*) \tau - (1 - \theta^*) t_p \right]$$  (27)

Finally, applying the definition of $\rho = \frac{t_l^*}{t_h^* + t_l^*}$ yields the result:

$$\rho \theta^L + (1 - \rho) \theta^H = \theta^* + (1 - \theta^*) \frac{t_p}{\tau}$$

Discussion of Proposition 3

Proposition 3 demonstrates that the government’s choice of tax salience raises special considerations in the context of Pigouian taxation. Intuitively, taxes on externality-generating
activities can only internalize the social costs of those activities to the extent that decision-makers account for the existence of the tax when choosing their behavior. When a Pigouvian tax increases social welfare by discouraging taxpayers from engaging in a particular activity, the government will face an additional efficiency cost to reducing the salience of that tax. Conversely, when the taxed activity generates positive externalities, the efficiency benefits to relying on low-salience taxes are greater than would otherwise be the case.\textsuperscript{28}

To illustrate, suppose that $\theta^H = 1$, $\theta^L = 0$, and $\theta^* = 0$ for some good $x$. Suppose this good generates a negative externality such that the optimal Pigouvian tax is given by $t_p > 0$. Equations 26 and 27 imply that the optimal policy is to set $t_h = t_p$ and $t_l = \tau - t_p$. When $\theta^H < 1$, the optimal size of the high-salience tax will need to be scaled up to fully internalize the externality with respect to taxpayers' consumption decisions: $t_h = t_p/\theta^H$, $t_l = \tau - t_p/\theta^H$.

Finally, when $\theta^H = 1$, $\theta^L = 0$, and $\theta^* \in (0, 1)$, the optimal policy is to set $t_h$ as a weighted average between raising revenue and correcting the externality, $t_h = \theta^* \tau + (1 - \theta^*) t_p$, and $t_l = (1 - \theta^*) \tau - (1 - \theta^*) t_p$.

Proposition 3 has important practical implications for policymakers concerned with bringing about behavioral changes on the part of taxpayers. For example, to reduce population weight, a number of states levy sales taxes on soda and/or candy while exempting other food purchases from the sales tax base. Although this approach may increase the true relative price of unhealthful foods, the analysis here suggests that such taxes would be more likely to generate the intended behavioral effects if they were designed in more salient ways.

6. Conclusion

A long literature within public finance considers how to mitigate the excess burden of distortionary taxation. Motivated by new empirical findings that a tax’s salience affects consumer behavior, this paper shows that seemingly-mundane choices about tax design may have surprisingly large effects on consumer welfare; making decisions about tax design in a strategic way can ease the burden of distortionary taxation. More generally, the results illustrate that

\textsuperscript{28}This intuition is also discussed in Finkelstein (2009).
careful attention to the biases that characterize individual decision-making may offer policymakers unexplored possibilities for improving consumer welfare through the manipulation of commonly-available (but frequently overlooked) tools.

One feature of my results that may be surprising is that the optimal policy requires the government to design its taxes in a way that drives consumers to make mistakes. That is, even when the government has access to a fully-salient tax ($\theta_h = 1$), Corollary 1.1 implies that taxpayers are actually better off when the government also employs a tax that causes them to make mistakes. This result can be readily understood as an example of the Theory of the Second Best (Lipsey and Lancaster 1956). That is, the government’s need to raise revenue through a commodity tax generates a distortion that pushes social welfare away from the first-best welfare outcome. Consequently, by introducing a new distortion – taxpayer deviations from optimal decision-making – policymakers can actually increase social welfare.

On a more practical level, the results here suggest that governments should typically avoid relying exclusively on fully-salient taxes (unless the purpose of the tax is solely to reduce consumption of the taxed good). Instead, the optimal combination of high- and low-salience taxes should depend on the nature of demand for the taxed good, in particular the ratio of the income and substitution effects. Similarly, policymakers should be skeptical of calls for tax-inclusive price regulations of the type common outside of the United States. Although Section 4 showed that the first-best welfare outcome may not be attainable when tax salience is endogenous and revenue demands are large, the results described in that section provide important guidance to policymakers about making incremental efficiency improvements.

Along the same lines, the results highlight the efficiency-enhancing potential of a tax instrument that is sustainably low-salience, i.e. a tax that remains low-salience even when it is levied at high rates. As such, an important takeaway of this paper is the desirability of new research into the factors that shape consumers attentiveness to a tax, and in particular, to the conditions that determine whether consumers remain inattentive as the tax rate increases. In addition to underscoring the need for more emphasis on tax design in public finance research, my results also suggest the need to reconsider accepted intuitions in the field regarding the
proper role of commodity taxation. For example, the Atkinson-Stiglitz theorem stands for the 
proposition that linear commodity taxes are undesirable, apart from special cases. However, 
when the government has multiple options for designing commodities taxes, and the options 
differ in their salience, my results suggest some role for commodity taxes (at least when the 
income tax is fully-salient). Such issues deserve further exploration. 29

Finally, by focusing on the case of a representative consumer, I have ignored distributional 
consequences associated with the choice between high- and low-salience taxes. In practice, 
decision-makers may exhibit behavioral biases in ways that correlate with policy-relevant 
characteristics (e.g. wealth), and such patterns can have important implications for the design 
of policy (Mullainathan and Shafir 2013; Shah, Mullainathan and Shafir 2012). Along these 
lines, tax salience may affect the distribution of a tax’s burden in several important ways. 
For example, when consumers are heterogeneous in their attentiveness to low-salience taxes, 
setting the salience of a tax affects the distribution of the tax’s burden between attentive 
and inattentive consumers. This case is explored by Goldin and Homanoff (2013), which 
shows how governments can manipulate tax salience to reduce commodity tax regressivity 
when high- and low-income consumers differ in their attentiveness to low-salience taxes. 
Additionally, tax salience affects the incidence of a tax between consumers and producers. 
In particular, Chetty et al. (2009) show that reducing tax salience can increase the fraction 
of the tax passed on to consumers. As in other contexts, the ultimate distributional effects 
of the choice between high- and low-salience taxes depend upon the supply and demand for 
the factors employed in production of the taxed good.

29Another question that merits further exploration concerns the optimal combination of high- and low-
salience taxes in the context of multiple taxed-goods. In order to investigate issues of tax design, I have 
focused on the case of a single taxed good and treated as exogenous the amount of revenue required to be 
rased from taxes on that good. Because the government cannot realistically raise all of its revenue from 
commodity taxes on a single good, issues of salience may have important interactions with the question of 
how to apportion taxes between goods, especially when the salience of a tax is a function of its size.
References


Appendix A

Proposition A.1

Suppose that utility is additively separable in x and y. Let ρ denote the fraction of taxes on x that are low-salience: \( ρ \equiv \frac{u}{t_b + t_i} \). Then under (A1) - (A3), a revenue-neutral shift towards the low-salience tax is desirable if and only if \( ρθ^L + (1 - ρ)θ^H > θ^* \), where \( θ^* \) is defined as in Proposition 2.

Proof of Proposition A.1

Under additive separability, utility is given by \( U(x, y) = u(x) + v(y) \). We can therefore rewrite (9) as:

\[
\left. \frac{dv}{dt_i} \right|_{t_i^0} \geq 0 \iff u'(x) - p v'(y) \geq 0
\]

(28)

Let \((x_0, y_0)\) denote the consumption bundle the consumer would choose absent any taxes, \(x_0 = x(0, 0, p), y_0 = y(0, 0, p)\). Because the after-tax price of \( x \) is fully-salient when there are no taxes, (A2) implies that \((x_0, y_0)\) satisfies the first order condition for an interior solution
from the consumer’s standard utility maximization problem:

\[ u'(x_0) = p v'(y_0) \quad (29) \]

First-order Taylor approximations of \( u'(.) \) and \( v'(.) \) around \( x_0 \) and \( y_0 \) (respectively) yield

\[ u'(x) \approx u'(x_0) + u''(x_0)(x - x_0) \quad (30) \]

\[ v'(y) \approx v'(y_0) + v''(y_0)(y - y_0) \quad (31) \]

Similarly, we can approximate\(^{30}\)

\[ x - x_0 \approx \frac{\partial x}{\partial p} \left( \theta^H t_h + \theta^L t_l \right) \]

(32)

Substituting (30), (31), and (32) into (28), and utilizing the individual budget constraint (1) and (29), we obtain:

\[ \frac{dv}{dt_l}|_{R_0} \geq 0 \iff \gamma_0 \left( \theta^H t_h + \theta^L t_l \right) \frac{\partial x}{\partial p} + p(t_h + t_l) x v''(y_0) \geq 0 \quad (33) \]

where \( \gamma_0 = u''(x_0) + p^2 v''(y_0) \).

To prove the result, we will need to express (33) in more familiar terms. Implicitly differentiating the consumer’s first order condition and budget constraint at \((x_0, y_0)\) yields

\[ \frac{\partial x}{\partial I}(0,0,p) = \frac{p v''(y_0)}{\gamma_0} \quad \text{and} \quad \frac{\partial \bar{x}}{\partial p}(0,0,p) = \frac{v'(y_0)}{\gamma_0}, \]

where \( \bar{x} \) represents compensated (Hicksian) demand for \( x \). Substituting these results into (33) and rewriting the equation in terms of elasticities yields:

\[ \frac{dv}{dt_l}|_{R_0} \geq 0 \iff \frac{t_h \theta^H + t_l \theta^L}{t_h + t_l} \geq \frac{\omega_x \eta_{x,I}}{\varepsilon_{x,p}} \quad (34) \]

where \( \omega_x = \frac{p x}{I} \) denotes the budget share of expenditures on \( x \), \( \eta_{x,I} = \frac{\partial x}{\partial I} \bigg|_{x_0} \) denotes the income-elasticity of \( x \), and \( \varepsilon_{x,p} = -\frac{\partial^2 x}{\partial p^2} \bigg|_{x_0} \) denotes the own-price elasticity of \( x \), defined to be

\(^{30}\)Note that \( \frac{\partial x}{\partial p} \) is evaluated at \((t_h, t_l, p) = (0, 0, p)\). As is standard in this literature (e.g. Auerbach (1985) and Chetty et al. (2009)), I ignore second-order terms in this approximation.
positive. Note that all of these quantities are evaluated at \((t_h, t_l) = (0, 0)\). Finally, rewriting (34) in terms of the definition of \(\theta^* = \frac{\omega_{x, y} \psi_{x, p}}{x_{x, p}}\) and \(\rho = \frac{t_l}{t_h + t_l}\) yields the intended result.

**Figure 1**