Optimal Tax Salience

Jacob Goldin*

August 21, 2015

Abstract

Recent empirical work finds that consumers under-account for commodity taxes when the after-tax price is not prominent. I investigate how policymakers may utilize such "low-salience" taxes to promote welfare. The optimal combination of high- and low-salience taxes balances two competing effects: low-salience taxes dampen distortionary substitution but cause consumers to misallocate their budgets. Using a stylized model, I show the availability of taxes with differing salience provides an extra degree of freedom that can be used to implement the first-best welfare outcome. I characterize the optimal policy and derive a formula for incremental adjustments when the first-best is unattainable.

*Princeton University. For helpful comments, I am grateful to Constantine Angyridis, Raj Chetty, Jonah Gelbach, Mikhail Golosov, Joseph Grundfest, Nikolaj Harmon, Tatiana Homonoff, Louis Kaplow, Alvin Klevorick, Nicholas Lawson, Yair Listokin, David Lee, Daniel Markovits, Alex Mas, Daniel Reck, and Dean Spears. All errors are my own.

1. Introduction

Optimal commodity taxation is a classic subject in public finance. Most research studies how governments should levy taxes across distinct goods to promote social welfare when lump-sum taxes are unavailable. In contrast, questions relating to tax design have not received the same degree of theoretical attention.[1]

Recent empirical work suggests a need to reconsider this emphasis. A series of findings suggests that the *salience* of a tax has important effects on consumer behavior: the less prominent the after-tax price of a good, the less consumers respond to changes in the tax on that good.[2]

Such findings suggest an additional margin through which governments can shape the behavioral effects of a tax. Although policymakers typically lack perfect control over a tax's salience, they frequently face a choice between relying on high- and low-salience ways of raising revenue. For example, policymakers can manipulate the salience of a commodity tax by choosing whether to include the tax in the displayed price of the taxed good or to add it on at the register when the consumer completes her purchase. Because the former is more salient than the latter, the government can alter the tax's salience by adjusting the degree to which it relies on the two tax designs.[3]

This paper studies the optimal salience of commodity taxes: how should a benevolent government choose between high- and low-salience taxes on a particular good to raise rev-

[1]By "tax design," I mean policy decisions relating to characteristics of a tax that do not directly enter into consumers' budget constraints. Two exceptions are Slemrod and Kopczuk (2002) and Krishna and Slemrod (2003).

[2]I use "salience" to refer to the prominence of the taxed good's tax-inclusive price. For example, 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 to the retailer is tax as opposed to the pre-tax price. For empirical research relating to commodity tax salience, see Chetty, Looney and Kroft (2009); Feldman and Ruffle (2015); Goldin and Homonoff (2013); Bradley and Feldman (2015). Krishna and Slemrod (2003) and McCaffery (1994) review earlier evidence on how tax design affects behavior.

[3]Policymakers 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. Such regulations are common in Europe but are rare in the United States. Similarly, governments may require tax-inclusive pricing for a particular good. For example, the Federal Trade Commission requires airlines to include taxes and other fees in the initial price displayed to consumers. Policymakers may also shape salience in other contexts: 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, 2014); and income tax payments may be collected from employees or automatically withheld (Jones, 2010).

enue? The analysis highlights two distinct mechanisms through which tax salience affects consumers' well-being. On the one hand, low-salience taxes dampen the type of 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 consumers to misallocate income among consumption goods. The government's choice between high- and low-salience taxes trades off between these competing effects.

In the standard model, the presence of an untaxed good causes optimal policy to diverge from the first-best; commodity taxes generate excess burden by distorting consumption decisions for taxed and untaxed goods. In contrast, when the government can control the salience of a tax, that flexibility provides an additional degree of freedom. I show that when the government can utilize two taxes on a single good that differ in their salience, it can employ those taxes in combination to achieve the first-best welfare outcome – even when one of the available goods cannot be taxed. The key insight is that by adjusting the balance between high- and low-salience taxes, the government can maintain a given level of revenue while causing taxpayers to vary their consumption of the taxed and untaxed good; in this way, taxpayers can be induced to choose the same allocation they would choose under a lump-sum tax (even though that allocation is privately sub-optimal given the taxes that are actually in place).

I next turn to characterizing the optimal combination of high- and low-salience taxes. Solving the government's problem yields an intuitive formula for the optimal policy, which highlights the link between optimal salience and the nature of demand for the good being taxed. Notably, the formula implies that 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.

In practice, adopting policies that are designed to exploit people's biases raises several important concerns. Although many of these, such as political transparency and credibility, are outside the scope of this paper, one that can undermine the results presented here is the possibility that taxpayers will become more attentive to low-salience taxes as the government increases its reliance on them – i.e., as the utility cost of neglecting the taxes grows larger. Before concluding, I consider an extension of the model to a setting in which the salience of a tax is endogenously related to the tax's size and derive conditions under which the first-best will be attainable. When the first-best is unattainable, I show how incremental adjustments in the balance between high- and low-salience taxes can still yield efficiency gains.

Despite the ubiquity of policy decisions that affect tax salience, the topic has received little theoretical attention. As Congdon, Kling and Mullainathan (2009) conclude in their review of the behavioral tax literature, "the theoretical literature has yet to yield the type of rules of thumb with respect to optimal tax salience that translate into practical policy recommendations." The research closest to the current analysis are Chetty, Looney and Kroft (2009), Chetty (2009), and Reck (2015). Those authors derive formulas for quantifying the excess burden of a tax that is less than fully salient but do not consider the implications of salience for optimal taxation. In addition, this paper is the first to consider the possibility of combining tax instruments that differ in their salience, and it is that possibility which drives the theoretical insights described here.

A number of influential papers have investigated how cognitive biases other than salience affect prescriptions for optimal tax policy (e.g., Liebman and Zeckhauser, 2004; O'Donoghue and Rabin, 2003). This literature evaluates the optimal level of a tax instrument conditional on taxpayers exhibiting an assumed behavioral bias. I build on this literature by studying a setting in which the government's choice of tax instrument controls the extent to which taxpayers exhibit the bias in the first place.

The remainder of the paper proceeds as follows. Section 2 develops the model and derives the main results – first graphically and then formally. Section 3 extends the model to account for the possibility that a tax's salience is endogenously related to its size. Section 4 concludes.

2. Model and Results

Society is composed of a representative taxpayer who divides her income I between goods x and y. Production of x is characterized by constant returns to scale technology so that its pre-tax price is fixed at marginal cost p. Good y is the numeraire. The taxpayer's utility depends on consumption of x and y:[4]

$$U = U(x, y) \tag{1}$$

U is concave and smooth with respect to both goods.

The government's objective is to maximize the representative taxpayer's utility while raising revenue R_0.

A. First-Best Welfare Outcome

Before turning to tax salience, it is helpful to characterize the first-best welfare outcome – i.e., what the government could achieve with access to a non-distorting tax. To derive this benchmark I will assume for purposes of this section that the government may levy a (fully-salient) lump-sum tax of size L.

When facing the lump-sum tax, the taxpayer's budget constraint is given by

$$p\,x + y = I - L \tag{2}$$

and her consumption satisfies the first-order condition associated with maximizing utility subject to this constraint:

[4] Expressing utility as a function of consumption is standard in public finance models but implies that other factors (such as tax salience) do not affect welfare apart from their effect on consumption. For example, an agent would violate the assumption if she preferred facing a register tax to a posted tax on political grounds, perhaps because the amount going to the government is more transparent under the former than the latter. If low-salience taxes do generate direct welfare costs to consumers the results presented here will overstate the benefits of low-salience taxes. However, as Chetty, Looney and Kroft (2007) show, even relatively small cognitive costs generate substantial under-reaction to a tax; consequently, omitting such costs from the model may not be as misleading as would otherwise be the case. In addition, note that not all psychic cost models are ruled out: suppose that accounting for a low-salience tax is associated with some cognitive cost, but because of that cost, the consumer rationally chooses to ignore the tax. This agent's utility function can be described by (1) because given her decision-making strategy, she does not suffer any direct utility cost when confronted with the tax.

$$U_x(x, y) = p\, U_y(x, y) \tag{3}$$

Because the revenue collected by a lump-sum tax of size L is simply L, the government's revenue constraint is satisfied if and only if

$$L = R_0 \tag{4}$$

Equations (2)-(4) pin down consumption under a lump-sum tax and hence characterize the first-best welfare outcome.

B. Tax Salience

Having characterized the first-best, I assume now that the government lacks access to a lump-sum tax and can only raise revenue through commodity taxes on x. Good y (the numeraire) is untaxed. The government has at its disposal two tax designs that it can levy on purchases of x: a high-salience tax t_h and a low-salience tax t_l. Both t_h and t_l are unit taxes. The taxpayer's budget constraint takes the form:

$$y + (p + t_h + t_l)\; x = I \tag{5}$$

Taxpayer Behavior

Taking income as fixed, demand for x and y can be written as a function of the two taxes and the pre-tax price of x: $x = x(p, t_h, t_l)$ and $y = y(p, t_h, t_l)$. To capture the empirical findings described in the introduction, I assume that the extent to which a tax affects consumer demand depends on the tax's salience. As in Chetty, Looney and Kroft (2009), I adopt a functional definition of tax salience.[5] For $i \in \{h, l\}$, the salience of a tax, θ_i, measures how taxpayers adjust their demand for the taxed good in response to a change in the tax relative

[5]This functional definition of salience corresponds to the common understanding of salience as "prominence" if the pre-tax price of a good is prominent to consumers and consumers fully account for the pre-tax price when making purchasing decisions. Equation (6) assumes a constant degree of under-reaction by consumers when a tax is less than fully-salient. I relax this assumption in Section 3.

to a change in the taxed good's pre-tax price:

$$\theta_h = \frac{\partial x / \partial t_h}{\partial x / \partial p} \qquad \theta_l = \frac{\partial x / \partial t_l}{\partial x / \partial p} \tag{6}$$

To illustrate the notation, a tax that appeared as part of the taxed good's posted price would be fully-salient ($\theta = 1$). In contrast, a tax to which consumers were entirely unresponsive would have $\theta = 0$. I assume that the two tax designs available to the government have differing (but individually-fixed) degrees of salience; the taxpayer is more responsive to changes in the high-salience tax than to changes in the low-salience tax:

$$0 \leq \theta_l < \theta_h \leq 1 \tag{7}$$

Whether Equation (7) is satisfied in a particular context is an empirical question. One common situation in which (7) will be satisfied is when the government has access to one commodity tax instrument that is less than fully salient ($\theta_l < 1$), such as a sales tax, and another that directly affects the posted price of the taxed good ($\theta_h = 1$), such as an excise tax. Equation (7) also imposes that the salience of the two tax instruments are between 0 and 1.

Finally, I assume throughout that $x + (t_h + t_l) \frac{\partial x}{\partial t_i} > 0$ for $i \in \{h, l\}$, which guarantees the government cannot raise additional tax revenue by reducing either of its taxes.

Behavioral Welfare Framework

In the standard model, demand for x and y correspond to the solution to the consumer's welfare maximization problem: $\text{MAX}_{x,y}\ U(x,y)\ \ s.t.\ \ y + (p + t_h + t_l)\ x = I$. Let $x^*(p, t_h, t_l)$ and $y^*(p, t_h, t_l)$ denote the quantities of x and y that maximize the taxpayer's utility subject to the budget constraint (as above, the income term in these functions is suppressed). It is straightforward to show that x^* and y^* depend only on the total after-tax price of the taxed good. Yet as discussed in Chetty, Looney and Kroft (2009), this result is inconsistent with the empirical evidence that consumer behavior depends in part on tax salience (rather than the size of the tax alone).[6] Consequently, the neoclassical revealed preferences approach to welfare analysis – which assumes rational decision-making by consumers – is inappropriate for analyzing policy decisions about tax salience.

Instead, I follow Chetty, Looney and Kroft (2009) by utilizing what Bernheim and Rangel (2009) refer to as a "refinement." Rather than assume that every decision the taxpayer makes reflects her true preferences, I assume the taxpayer behaves optimally when tax-inclusive prices are fully salient (e.g., when all taxes are included in the posted price)[7]

$$x(p + t_h + t_l, 0, 0) = x^*(p, t_h, t_l) \tag{8}$$

Government's Problem

As above, I consider the problem faced by a government seeking to maximize consumer welfare subject to a revenue constraint R_0. It will be convenient to express consumer welfare as a function of the government's choice of taxes:

$$V(t_h, t_l) = U\left(x(p, t_h, t_l)\,,\, y(p, t_h, t_l)\right) \tag{9}$$

Total government revenue R is also a function of the taxes: $R(t_h, t_l) = (t_h + t_l)\ x(p, t_h, t_l)$. The government's revenue constraint is therefore given by

$$(t_h + t_l)\ x(p, t_h, t_l) = R_0 \tag{10}$$

The government's problem is to choose the combination of t_h and t_l that solves

$$\text{MAX}_{t_h, t_l}\ V(t_h, t_l) \quad s.t. \quad (t_h + t_l)\, x(p, t_h, t_l) = R_0 \tag{11}$$

[6] That is, if consumers behave optimally, (1) and (5) and are inconsistent with (6) and (7).

[7] Equation (8) is a weaker version of the typical rationality assumption underlying the revealed preference approach to welfare analysis. Rather than assume that all of a decision-maker's choices reflect her true preferences, this approach imposes rationality only for the subset of choices made when taxes are fully salient. Because there are good reasons to be skeptical about the quality of choices made when taxes are less than fully-salient, this approach privileges the preferences revealed when those conditions are not present.

C. Graphical Illustration

This section provides graphical intuition for the main result of the paper: having access to two taxes with differing salience provides the government an extra degree of freedom that it can use to implement the first-best welfare outcome. The next section provides a formal proof.

Consider a stylized example, depicted in Figure 1. Suppose the government is choosing between a fully-salient tax ($\theta_h = 1$) and a tax to which consumers are entirely unresponsive ($\theta_l = 0$). The consumer's pre-tax budget constraint is given by the line AB, and pre-tax consumption (E_0) is characterized by the tangency of the consumer's indifference curve with AB. Because any feasible choice of taxes must raise revenue R_0, the taxpayer's consumption under any feasible tax combination will lie somewhere on the line CD, which is simply line AB shifted downwards by the vertical distance R_0.

To identify consumption under the first-best, consider a lump-sum tax of size R_0. Because lump-sum taxes do not affect relative prices, the budget constraint induced by the lump-sum tax is also given by line CD. Consumption under the first-best allocation (E_{LST}) is determined by the point at which the consumer's indifference curve is tangent to line CD.

Now, suppose the government relies solely on t_h for raising revenue. In that case, the consumer's budget constraint would pivot to line AF. Consumption under this tax (E_h) is the tangency point between line AF and the consumer's indifference curve. Note that the high salience tax generates excess burden by driving consumers to substitute away from the taxed good over and above the income effect of the tax, $x_h < x_{LST}$.

In contrast, if the government were to rely solely on t_l for raising revenue, demand for the taxed good would not change from its pre-tax level, inducing consumption at E_l. 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 income effect, $x_l > x_{LST}$.

Intuitively, because all feasible tax combinations induce consumption along the line CD, and because consumption under the lump-sum tax lies between the consumption induced by the high- and low-salience taxes (when either is imposed alone), the optimal policy lies somewhere between full reliance on t_h or t_l. That is, by shifting the balance between the high- and low-salience tax, the government can move consumption along CD until it reaches E_{LST}. In particular, suppose the government imposes the high salience tax at a level – call it t_h^* – that pivots the consumer's budget constraint to line AG, thereby inducing consumption at E_*, where demand for the taxed good is equal to demand for the taxed good under the first-best. Since E_* lies above line CD, this tax, on its own, fails to meet the government's revenue constraint. However, the government can combine t_h^* with a low-salience tax (t_l^*) to make up the additional revenue. And because $\theta_l = 0$, imposing the low-salience tax does not change the amount of x demanded by the taxpayer; it simply shifts consumption downwards. In this way, the government can combine t_h^* and t_l^* to induce E_{LST} – the same consumption that would be induced under a lump-sum tax.[8]

This simple example illustrates how the availability of multiple tax instruments that differ in their salience provides policymakers an additional degree of freedom with which to shift consumer demand while maintaining a desired level of revenue. The following sections formalize this intuition and explore the conditions that must be met for the first-best to be attainable.

D. Welfare Under Optimal Tax Salience

To derive the optimal policy, I begin with an arbitrary (feasible) combination of high- and low-salience taxes and consider the welfare consequences of (feasible) adjustments to the initial combination. The optimal policy corresponds to the combination of taxes for which no feasible adjustment would increase welfare.

Consider some combination of taxes (t_h, t_l) that satisfies the government's revenue constraint, $(t_h + t_l)\ x\,(p, t_h, t_l) = R_0$. The government may adjust t_h and t_l, but in order for the

[8]Note that although E_{LST} is the socially-optimal allocation, it is privately sub-optimal under t_h^* and t_l^*. That is, any individual taxpayer facing t_h^* and t_l^* would be (privately) better off consuming an allocation with less x and more y than E_{LST}.

combination to be feasible, it must adjust the taxes in such a way that the revenue constraint continues to hold. The feasible combinations of taxes can be found by totally differentiating the revenue constraint: $\left(x + (t_h + t_l)\frac{\partial x}{\partial t_h}\right) d\,t_h + \left(x + (t_h + t_l)\frac{\partial x}{\partial t_l}\right) d\,t_l = 0$. With (6), this yields the change in the high-salience tax associated with a small increase in the low-salience tax such that the overall policy change is revenue-neutral.

$$\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0} = -\frac{\theta_l \frac{\partial x}{\partial p}(t_h + t_l) + x}{\theta_h \frac{\partial x}{\partial p}(t_h + t_l) + x} \tag{12}$$

In words, a \$1 increase in the low-salience tax accommodates a revenue-neutral reduction in the high-salience tax of $\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0}$ dollars. Note that (7) implies $\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0} < -1$.[9] Intuitively, because taxpayers adjust their demand more in response to changes in the high-salience tax, a \$1 increase in the low-salience tax accommodates a revenue-neutral reduction in the high-salience tax of more than \$1.

Define a revenue-neutral shift towards the low-salience tax as a marginal increase in t_l along with the corresponding reduction in t_h needed to maintain revenue neutrality. The following result describes the welfare effect of this policy change.

Lemma 1 *The welfare effect of a revenue-neutral shift towards the low-salience tax,* $\left.\frac{dV}{dt_l}\right|_{R_0}$*, is given by*

$$\left.\frac{dV}{dt_l}\right|_{R_0} = \left[\underbrace{-x\left(1 + \left.\frac{\partial t_h}{\partial t_l}\right|_{R_0}\right)}_{1} + \underbrace{\left(\theta_l + \left.\frac{\partial t_h}{\partial t_l}\right|_{R_0}\theta_h\right)}_{2}\underbrace{\frac{\partial x}{\partial p}}_{3}\underbrace{\left(\frac{U_x}{U_y} - (p + t_h + t_{.l})\right)}_{4}\right] U_y(x, y)$$

Proof Totally differentiating the consumer's budget constraint (5) yields

$$\frac{\partial y}{\partial t_i} = -x - (p + t_h + t_l)\frac{\partial x}{\partial t_i} \text{ for } i \in \{h, l\} \tag{13}$$

In addition, totally differentiating the consumer's welfare (9) yields:

[9] In particular, this follows from $\theta_h > \theta_l$ along with the maintained assumption that $x + (t_h + t_l)\frac{\partial x}{\partial t_i} > 0$ for $i \in \{h, l\}$.

$$\left.\frac{dV}{dt_l}\right|_{R_0} = U_x(x,y)\,\frac{\partial x}{\partial t_l} + U_x(x,y)\,\frac{\partial x}{\partial t_h}\,\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0} + U_y(x,y)\,\frac{\partial y}{\partial t_l} + U_y(x,y)\,\frac{\partial y}{\partial t_h}\,\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0} \tag{14}$$

Substituting (6) and (13) into (14) yields the result. ■

The expression in Lemma 1 is the sum of two intuitive components. Term 1 captures the welfare effect from the salience shift on the taxpayer's purchasing power. The net change in taxes on x is given by $\left.\frac{d(t_h+t_l)}{dt_l}\right|_{R_0} = 1 + \left.\frac{\partial t_h}{\partial t_l}\right|_{R_0}$. Scaling that price change by the taxpayer's consumption of x yields the change in purchasing power. Term 1 is positive because, as explained above, $\theta_l < \theta_h$ guarantees $\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0} < -1$. Intuitively, because a revenue-neutral shift towards the low-salience tax accommodates a net reduction in taxes on x, there is a corresponding increase in the taxpayer's purchasing power.

The second piece of Lemma 1 (the product of Terms 2, 3, and 4) captures the welfare loss from optimization errors. To interpret Term 2, it is helpful to first define the *price-equivalent tax*, p_θ, to be the magnitude of the pre-tax price change that would induce the same change in demand for x as imposing t_l and t_h, $p_\theta = \theta_l t_l + \theta_h t_h$. Loosely speaking, this quantity reflects the after-tax price of x that is perceived by consumers. Term 2 is equal to the change in the price-equivalent tax induced by the shift, $\left.\frac{dp_\theta}{dt_l}\right|_{R_0}$, i.e., the change in the price of x as perceived by taxpayers. Term 3 maps the price-equivalent tax change into behavior; the product of Terms 2 and 3 reflects the increase in consumption of x induced by the policy shift. Finally, Term 4 maps the change in consumption into welfare. Note that the first-order condition associated with the taxpayer's welfare maximization problem implies that Term 4 equals 0; thus if the taxpayer's behavior were optimal, a marginal increase in consumption of x would have no effect on welfare. In contrast, when taxpayers misperceive low-salience taxes, Term 4 is negative because the marginal utility of expenditures on y exceeds the marginal utility of expenditures on x (consumption of x is sub-optimally high). Taken as a whole, the product of Terms 2, 3, and 4 capture the fact that shifting towards the low-salience tax causes taxpayers to incur welfare losses by over-consuming the taxed good.

Lemma 1 highlights the tension that characterizes the optimal salience problem: increasing the government's reliance on low-salience taxes reduces total taxes on x (raising consumers' purchasing power), but induces consumers to deviate further from their optimal consumption bundle. Note that were $\theta_h = \theta_l$ (in violation of Assumption (7)), both pieces of Lemma 1 would be equal to zero; intuitively, when both of the available tax instruments have the same salience, shifting between them does not affect welfare.

Under the optimal policy, no (feasible) shift in taxes generates an improvement in welfare, $\left.\frac{dV}{dt_l}\right|_{R_0} = 0$. This condition allows us to characterize the first-order condition to the government's problem.

Lemma 2 *The optimal combination of high- and low-salience taxes induces the taxpayer to consume values of x and y that satisfy*

$$U_x(x, y) - p\,U_y(x, y) = 0$$

Proof At the optimum, no feasible shift in taxes generates an improvement in welfare, $\left.\frac{dV}{dt_l}\right|_{R_0} = 0$. Setting the expression in Lemma 1 equal to zero and substituting in the expression for $\left.\frac{\partial t_h}{\partial t_l}\right|_{R_0}$ from (12) yields:

$$-x\left(1 - \frac{\theta_l \frac{\partial x}{\partial p}(t_h + t_l) + x}{\theta_h \frac{\partial x}{\partial p}(t_h + t_l) + x}\right) + \left(\theta_l - \theta_h \frac{\theta_l \frac{\partial x}{\partial p}(t_h + t_l) + x}{\theta_h \frac{\partial x}{\partial p}(t_h + t_l) + x}\right) \frac{\partial x}{\partial p} \left(\frac{U_x}{U_y} - (p + t_h + t_l)\right) = 0$$

or, after simplifying:

$$(\theta_h - \theta_l)\left(\frac{U_x}{U_y} - p\right) = 0 \qquad (15)$$

Assumption (7) guarantees $\theta_h - \theta_l \neq 0$; dividing (15) by that quantity yields the result.■

Comparing the conditions that characterize consumption under the lump-sum tax with the conditions that characterize consumption under the optimal combination of high- and low-salience taxes yields the main result of this section.

Proposition 1 *The optimal combination of high- and low-salience taxes achieves the first-best welfare outcome.*

Proof Consumption under the lump-sum tax is determined by (2), (3), and (4). Consumption under the optimal combination of high- and low-salience taxes is determined by (5), (10), and Lemma 2. Equations (2) and (4) imply $p\,x + y = I - R_0$, which is also implied by (5) and (10). In addition, (3) is identical to Lemma 2. Hence, consumption under the lump-sum tax is equal to consumption under the optimal combination of t_h and t_l. Equation (1) guarantees utility depends only on consumption; hence welfare under the two policies is the same as well. ■

Proposition 1 demonstrates that policy control over tax salience provides the government an extra degree of freedom with which to implement the first-best, even when one of the available goods cannot be taxed. As illustrated in Section 2.C, the basic intuition is that the first-best level of demand for the taxed good can be induced by different combinations of the tax instruments, but each combination yields a different amount of revenue. Implementing the first-best thus requires identifying which of these combinations yields sufficient revenue to meet the government's revenue constraint.[10]

E. Characterizing the Optimal Salience of a Tax

This section investigates what combination of high- and low-salience taxes are required to implement the optimal policy. I first derive a formula for the optimal policy by generalizing the graphical approach described in Section 2.C. I then relate the optimal degree of salience to observable elasticities, which sheds light on the economic forces that shape the optimal policy.

To combine high- and low-salience taxes in a manner that implements the first-best, consider the following approach. First, set $t_h = \overline{t_h}$, where $\overline{t_h}$ is defined as the level of the high-salience tax that, when employed without any other taxes, induces the first-best quantity of consumption of x. In Figure 1, $\bar{t}_h$ corresponds to t_h^*.[11] Writing demand for x as a function

[10] It is straightforward to extend this result to settings in which the taxed good generates a consumption externality. When the externality is negative, the optimal policy is for the government to rely more heavily on the high salience tax because the first-best entails less consumption of the taxed good.

[11] In Figure 1, $\bar{t}_h$ was not only the value of the high salience tax that induced the first-best consumption of x, it also happened to be the high salience tax's optimal value. More generally, when $\theta_l \neq 0$ or when $\theta_h < 1$, $\bar{t}_h$ may diverge from the optimal high salience tax, as described below.

of t_h, t_l, and after-tax income, the first-best can be written as the value of x induced by a lump-sum tax of size R_0, $x_{LST} \equiv x\,(0,0,I-R_0)$. Consequently, $\overline{t_h}$ is implicitly defined by $x\,(\overline{t_h},0,I) \equiv x\,(0,0,I-R_0)$.

Although $\overline{t_h}$ induces the first-best level of consumption of x, the revenue raised by $\overline{t_h}$ will not in general be equal to R_0. However, the difference in salience between the available tax instruments permit the government 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 combining increases in t_l with reductions in t_h, the government can increase revenue without causing individuals to substitute away from the taxed good. To see this, note that totally differentiating x (while holding after-tax income fixed) yields:

$$dx = \frac{\partial x}{\partial t_l}\partial t_l + \frac{\partial x}{\partial t_h}\partial t_h = \theta_l \frac{\partial x}{\partial p}\partial t_l + \theta_h \frac{\partial x}{\partial p}\partial t_h$$

It follows that movements along the line $\frac{\partial t_l}{\partial t_h}|_{x_{LST}} = -\frac{\theta_h}{\theta_l}$ do not affect the taxpayer's demand for x. On the other hand, movements along this line *do* affect the amount of revenue that is raised: $\frac{dR}{dt_h}|_{x_{LST}} = (t_l+t_h)\,\frac{\partial x}{\partial t_h}|_{x_{LST}} + x_{LST}\,\left(1+\frac{\partial t_l}{\partial t_h}|_{X_{LST}}\right)$. Because $\frac{\partial x}{\partial t_h}|_{x_{LST}} = 0$, we have:

$$\frac{dR}{dt_h}|_{x_{LST}} = -x_{LST}\,\left(\frac{\theta_h-\theta_l}{\theta_l}\right)$$

Thus for each \$1 reduction in t_h, t_l may be increased by $\frac{\theta_h}{\theta_l}$ dollars without causing consumption of x to depart from x_{LST}. At the same time, this policy change raises revenue in the amount of $x_{LST}\left(\frac{\theta_h-\theta_l}{\theta_l}\right)$ dollars.

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

$$\delta^* = (\tau - \overline{t_h})\left(\frac{\theta_l}{\theta_h-\theta_l}\right) \tag{16}$$

where $\tau \equiv \frac{R_0}{x_{LST}}$.

Using (16), we can now solve for $t_h^* = \overline{t}_h - \delta^*$ and $t_l^* = \frac{\theta_h}{\theta_l}\,\delta^*$, the values of the taxes that induce first-best consumption while satisfying the revenue constraint.

$$t_h^* \equiv \overline{t_h} - \delta^* = \left(\frac{\theta_h \overline{t_h} - \theta_l \tau}{\theta_h - \theta_l}\right) \tag{17}$$

$$t_l^* \equiv \frac{\theta_h}{\theta_l}\,\delta^* = (\tau - \overline{t_h})\left(\frac{\theta_l}{\theta_h - \theta_l}\right) \tag{18}$$

Equations (17) and (18) allow one to implement the first-best solution given knowledge of x_{LST} and $\overline{t_h}$. To implement the optimum when these quantities are not known, and to better understand the mechanisms at work, it is helpful to express t_h^* and t_l^* as functions of more familiar quantities. Let $\eta_{x,I} = \frac{\partial x}{\partial I}\frac{I}{x}$ denote the income-elasticity of x, $\omega_x = \frac{p\,x}{I}$ the budget share of expenditures on x, and $\varepsilon_{x,p} = -\frac{\partial x}{\partial p}\frac{p}{x}$ the own-price elasticity of x (defined to be positive), where each quantity is evaluated at the no-tax baseline.

Define

$$\theta^* = \frac{\eta_{x,I}\,\omega_x}{\varepsilon_{x,p}}$$

The numerator of θ^* represents the income effect associated with a price increase on x – the elasticity corresponding to the slope of the Engel curve through the no-tax optimum. The denominator of θ^* represents the combined income and substitution effects, $\varepsilon_{x,p} = \tilde{\varepsilon}_{x,p} + \omega_x\,\eta_{x,I}$, where $\tilde{\varepsilon}_{x,p}$ denotes the compensated (Hicksian) own-price elasticity of demand. By scaling the income effect by the combined income and substitution effects, a tax with salience θ^*induces taxpayers to adjust their consumption of x as if there were no substitution effect, but to still account for the tax's income effect (as they would under a lump-sum tax).

The following proposition shows that θ^* describes the optimal 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 exactly equal to θ^*, the optimal policy is to rely on that instrument exclusively. When no such tax is available, the government may replicate the welfare effects of a tax with optimal salience by combining the tax instruments that are available so that the weighted average replicates a single tax with salience θ^*.

Proposition 2 *Let ρ denote the fraction of taxes on* x *that are low-salience:* $\rho \equiv \frac{t_l}{t_h+t_l}$. *Let* $\theta^* = \frac{\eta_{x,I}\,\omega_x}{\varepsilon_{x,p}}$, *where each quantity is evaluated at the no-tax baseline. Then the optimal combination of high- and low-salience taxes is given by the value of ρ that solves*

$$\rho\theta_l + (1-\rho)\theta_h = \theta^*$$

Proof By the 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:

$$\overline{t_h}\theta_h \frac{\partial x}{\partial p} \approx -R_0 \frac{\partial x}{\partial I} \tag{19}$$

Using the definitions of $\eta_{x,I}$, ω_x, $\varepsilon_{x,p}$, and τ, it is straightforward to rewrite (19) as

$$\overline{t_h} \approx \left(\frac{\eta_{x,I}\,\omega_x}{\varepsilon_{x,p}}\right)\frac{\tau}{\theta_h} \tag{20}$$

Substituting (20) into (17) and (18) yields an expression for the optimal taxes in terms of θ^*.

$$t_h^* \approx \frac{\tau}{\theta_h - \theta_l}\ (\theta^* - \theta_l) \tag{21}$$

$$t_l^* \approx \frac{\tau}{\theta_h - \theta_l}\ (\theta_h - \theta^*) \tag{22}$$

Finally, noting that $t_h^* + t_l^* = \tau$, we can rewrite (22) in terms of ρ to obtain the result. ■

Proposition 2 yields a number of important insights. First, as discussed above, θ^* represents the ratio of the tax's income effect to its combined income and substitution effects; a tax with salience θ^* therefore induces taxpayers to set their consumption as if there was no substitution effect but only an income effect associated with the tax. In other words, introducing a new tax with salience θ^* to raise a marginal amount of revenue induces taxpayers to reduce their demand for x by following the Engel curve through the no-tax optimum – thereby replicating (locally) the behavioral effects of a lump-sum tax.[12]

[12] To see this formally, write $\theta^* = -x\frac{\partial x/\partial I}{\partial x/\partial p}$ so that a tax t with salience θ^* induces a change in demand of $\frac{\partial x}{\partial t} = \theta^* \frac{\partial x}{\partial p} = -x\frac{\partial x}{\partial I}$. The effect on x of raising a marginal amount of revenue using t is given by $\frac{\partial x}{\partial R} = \frac{\partial x/\partial t}{\partial R/\partial t} = \frac{-x\partial x/\partial I}{x - t\partial x/\partial t} = -\frac{\partial x}{\partial I}$ at the no-tax baseline $t = 0$. Additionally, the effect on x of raising a marginal

Second, for normal goods, $\theta^* \in [0, 1)$.[13] This implies that even when the government has access to fully salient tax ($\theta_h = 1$), the optimal policy is not to rely on it exclusively (i.e., $\rho > 0$) despite the fact that doing so would eliminate any mistakes by taxpayers. This result can be understood as an application of the Theory of the Second Best (Lipsey and Lancaster, 1956-67). That is, the government's need to raise revenue through a commodity tax creates 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 increase social welfare.

Third, Proposition 2 shows that the optimal combination of high- and low-salience instruments depends upon the nature of demand for the good being taxed. In particular, θ^* is declining in $\tilde{\varepsilon}$. Intuitively, the excess burden associated with a tax depends on the compensated elasticity of the taxed good (Auerbach, 1985). The greater is $\tilde{\varepsilon}$, the larger the welfare gains from reducing the consumer substitution that is typically associated with commodity taxes in the presence of an untaxed good. Additionally, θ^* is increasing in $\eta_{x,I}$ – the income elasticity associated with the taxed good. This is because the welfare cost of the budgeting mistake is larger for goods with higher income elasticities – neglecting the reduction in purchasing power caused by the tax leaves a taxpayer further from the amount of x she would consume at her (private) optimum. Note that $\theta^* = 0$ if and only if demand for the taxed good is entirely insensitive to income.[14]

Finally, Proposition 2 highlights the conditions under which subsidies will be required to implement the first-best. In particular, the optimal value of both tax instruments will be

amount of revenue using lump-sum tax L is also given by $\frac{\partial x}{\partial R} = \frac{\frac{\partial x}{\partial L}}{\frac{\partial R}{\partial L}} = -\frac{\partial x}{\partial I}$.

[13] Mechanically, this follows from the fact that $\tilde{\varepsilon}_{x,p} > 0$ when consumers behave optimally, and (8) guarantees that consumers behave optimally at the no-tax baseline, where $\tilde{\varepsilon}$ is evaluated.

[14] 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 t^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 $\eta_{x,I} = 0$, this response exactly matches how a fully-optimizing agent would respond to the tax. Because the optimal choice of x does not depend on income, the consumer has nothing to gain by reconsidering her consumption of x after a decline in income. In contrast, when $\eta_{x,I} > 0$, 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, Looney and Kroft (2009) for a closely related discussion.

non-negative if and only if $\theta^* \in [\theta_l, \theta_h]$.[15] When $\theta^* < \theta_l$, Equations (21) and (22) show that implementing the first-best requires utilizing a a low-salience tax in conjunction with high-salience subsidy.[16] When subsidies are unavailable and $\theta^* \notin [\theta_l, \theta_h]$, 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 θ^*.

3. Optimal Policy When Salience is Endogenous

So far I have assumed that the degree of salience associated with the available tax instruments is fixed and exogenous to the model. In practice, it may be that a tax's salience depends in part on its size. For example, some bounded rationality models of decision-making imply that consumers will pay more attention to larger taxes because the utility costs of neglecting a large tax are greater than those from neglecting a small tax (Chetty, Looney and Kroft, 2007; Reck, 2015).[17] When taxpayers behave in this way, the salience of the tax will be increasing in the tax's size and the results from previous sections may not hold.

Feasibility of the First-Best

This section derives conditions for whether the first-best is feasible in settings where tax salience is endogeneously related to the size of the tax. Suppose the government has two tax instruments available to it, a high-salience tax with θ_h fixed at 1 (such as an excise tax) and

[15]Mechanically, this follows from (21) and (22). 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 $\theta_h = 1$.

[16]This assumes the salience of a tax instrument is identical to the salience of a similarly-designed subsidy. In at least some contexts, tax subsidies may have higher salience than similarly-designed taxes (Feldman and Ruffle, 2015). Additionally, when subsidies are required to reach the first-best, (21) highlights the factors that shape how large the subsidy must be. First, when R_0 is large and the amount of x consumed under the first-best policy (x_{LST}) is small, τ 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 large. In the extreme case in which $\theta_h = \theta_l$, the required subsidy would be infinitely large and the result would not hold.

[17]On the other hand, researchers have documented behavioral biases even in decision-making contexts where the stakes are large, such as retirement savings decisions (Beshears et al., 2009), high-interest borrowing (Bertrand and Morse, 2011), labor supply decisions by earned income tax credit filers (Chetty and Saez, 2013), and property tax assessment appeals (Hayashi, 2014). With respect to commodity taxation, Feldman, Goldin and Homonoff (2015) finds no evidence that salience effects diminish as the tax rate increases in the context of a real-stakes laboratory experiment.

a low-salience tax for which the salience depends positively upon the tax's size, $\theta_l = \theta_l(t_l)$, $\frac{\partial \theta_l}{\partial t_l} > 0$. We can begin as before by setting t_h at the level necessary to induce consumers to consume x at the first-best quantity, $t_h = \overline{t_h}$. As before, consider a reduction in t_h along with an increase in t_l so that the net effect is for taxpayers to continue consuming x at x_{LST}.

Totally differentiating demand for x yields $\frac{\partial t_h}{\partial t_l}|_{x_{LST}} = -\theta_l(t_l)$. As a result, the additional revenue generated by an "x-neutral" increase in t_l is given by

$$\frac{dR}{dt_l}|_{x_{LST}} = \frac{\partial(t_l + t_h)}{\partial t_l}|_{x_{LST}}\, x_{LST} = (1 - \theta_l(t_l))\ x_{LST} \tag{23}$$

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 $\overline{t_h}\, x_{LST} + \int_0^{\hat{t_l}} \frac{\partial R}{\partial t_l}|_{x_{LST}}\ \partial t_l \geq R_0$, or, using (23):

$$\overline{t_h} + \int_0^{\hat{t_l}} (1 - \theta_l(t_l))\ \partial t_l \geq \tau \tag{24}$$

where $\tau = \frac{R_0}{x_{LST}}$.[18]

Finally, as before, $\overline{t_h}$ can be expressed in terms of familiar quantities by noting that $x(p, \overline{t_h}, 0, I) \equiv x(p, 0, 0, I - R_0)$. Taking first-order Taylor approximations around $x(p, 0, 0, I)$ implies $\overline{t_h}\frac{\partial x}{\partial p} \approx -R_0 \frac{\partial x}{\partial I}$. Writing this result in terms of elasticities yields $\overline{t_h} \approx \tau\, \theta^*$, where $\theta^*$is defined as in Proposition 2. Substituting this approximation into (24) implies that the optimal combination of high- and low-salience taxes achieves the first-best welfare outcome if and only if

$$\int_0^{\hat{t_l}} (1 - \theta_l(t_l))\ \partial t_l \geq \tau\ (1 - \theta^*) \tag{25}$$

Thus, when the salience of the available tax instruments is endogenous, determining

[18]One possibility, inconsistent with Equation (1)'s implication that welfare depends solely on consumption, 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, (24) 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-best because they are suffering the psychological costs associated with paying some attention to the low-salience tax. Note that to the extent that attentiveness to the taxes is (locally) stable, policymakers may still employ the incremental approach described later in this section for adjusting the balance between high- and low-salience taxes.

whether the first-best welfare outcome is feasible depends on three factors. First, because τ depends positively on R_0, the greater the revenue that must be raised from taxes on x, the harder it will be to attain the first-best. Second, it will be easier to attain the first-best when θ^* large, e.g., when demand for the taxed good is relatively income elastic. Intuitively, achieving the first-best in such cases requires relying less heavily on the low-salience tax, reducing the likelihood that consumers will become more attentive to it. Finally, whether the first-best can be achieved depends on the relationship between tax size and salience. The slower that $\theta_l(.)$ increases when the government increases its reliance on t_l, the more likely that the first-best will be feasible.

Local Improvements when the First-Best is Infeasible

Even when the salience of the available tax instruments increases too fast to achieve the first-best, the results here can shed light on whether incremental changes in the balance between high- and low-salience taxes is desirable. In particular, suppose that for the current values of t_h and t_l, θ_h and θ_l are such that $\rho\,\theta_h(t_h) + (1-\rho)\,\theta_l(t_l) > \theta^*$, where ρ and θ^* are defined as in Proposition 2. In such cases, it is straightforward to show that the welfare effect of an incremental revenue-neutral shift towards the low-salience tax will be positive, and vice versa when $\rho\,\theta_h(t_h) + (1-\rho)\,\theta_l(t_l) < \theta^*$. This claim is formalized in the Appendix.

Because of this, computing θ^* and identifying θ_l and θ_h at the current tax rates is sufficient to assess whether a small adjustment in salience will generate efficiency benefits. For example, if the government decides that it is going to raise total taxes on x by a small amount, this formula provides guidance for selecting which one of the available tax instruments should be increased. In contrast, if the planned tax increase is large, policymakers should be cautious of relying on this formula because the change in the magnitudes of the taxes could induce changes in their salience.

4. Conclusion

A long literature within public finance considers how to minimize the efficiency cost of distortionary taxation. Motivated by new empirical findings that a tax's salience affects consumer behavior, this paper explored how attention to salience can provide policymakers with an extra degree of freedom for reducing a commodity tax's excess burden. More generally, the results illustrate that careful attention to decision-making biases offer unexplored possibilities for improving consumer welfare through the manipulation of commonly-available (but frequently overlooked) policy tools.

Several limitations are important to keep in mind when interpreting the theoretical results presented here. Most importantly, the model abstracts from considerations that may shape the optimal degree of tax salience in the real world. For example, when agents are heterogeneous in the extent to which they respond to a given tax – that is, when a single tax instrument has different salience for different decision-makers – it will not in general be possible to achieve the first-best, at least when all agents must face the same tax instruments. Nonetheless, it would be straightforward to generalize the approach described in Section 3 for making incremental adjustments in salience to settings characterized by such heterogeneity. Along the same lines, by focusing on the case of a representative consumer, I have ignored distributional effects from the choice between high- and low-salience taxes. In reality, decision-makers may exhibit behavioral biases in ways that correlate with individual characteristics, and such patterns can have important implications for the design of policy. For example, when high- and low-income consumers differ in their attentiveness to low-salience taxes, governments can manipulate tax salience to reduce commodity tax regressivity (Goldin and Homonoff, 2013). Similarly, the salience of a tax may affect its incidence between consumers and producers (Chetty, Looney and Kroft, 2009).

Another limitation is that, contrary to what is assumed here, the government's objective function might seek to avoid policies that would induce its citizens to make mistakes. If so, policymakers may not wish to implement the optimal combination of high- and low-salience taxes because doing so would induce taxpayers to (accidentally) depart from the allocation that would be privately optimal for them to consume. For further discussion of such issues, refer to Galle (2009), Gamage and Shanske (2011), and Goldin (2012).

Finally, the results highlight several promising avenues for future research. The first 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 will remain inattentive as the size of the tax increases. Such research would be beneficial given the efficiency-enhancing potential of a tax instrument that is "sustainably" low-salience – i.e., that remains low-salience even when levied at high rates – as discussed in Section 3. Second, the results 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 commodity taxes are undesirable in the presence of a non-linear income tax, apart from special cases. However, when the government has multiple options for designing commodities taxes, and the options differ in their salience, the results here suggest that some role for commodity taxes may be optimal (at least when the income tax is fully-salient). Finally, it may be worthwhile to consider the implications of salience for the optimal allocation of taxes across commodities. The canonical Ramsey rule suggests levying taxes based on the elasticity of consumers' demand for the taxed goods; the results here suggest the optimal policy depends on whether observed elasticities stem from inelastic preferences or from the imposed taxes having low salience.[19]

References

Auerbach, Alan. 1985. "The Theory of Excess Burden and Optimal Taxation:." *Handbook of Public Economics*, 1.

Bernheim, B Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal*

[19]A new paper fleshing out this notion (and many others) is Farhi and Gabaix (2015). As noted by those authors, their framework provides a promising avenue for extending the optimal salience results derived here to settings with multiple goods and heterogeneous agents.

of Economics, 124(1): 51–104.

Bertrand, Marianne, and Adair Morse. 2011. "Information Disclosure, Cognitive BiBias and Payday Borrowing." *Journal of FInance*, 66.

Beshears, John, James J Choi, David Laibson, and Brigitte C Madrian. 2009. "The importance of default options for retirement saving outcomes: Evidence from the United States." In *Social security policy in a changing environment*. 167–195. University of Chicago Press.

Bradley, Sebastien, and Naomi Feldman. 2015. "Hidden Baggage: Air Traveler Responses to Add-On Taxes and Fees." *Working Paper*.

Chetty, Raj. 2009. "The Simple Economics of Salience and Taxation:." *NBER Working Paper 15246*.

Chetty, Raj, Adam Looney, and Kory Kroft. 2007. "Salience and Taxation: Theory and Evidence." *NBER Working Paper 13330*, , (4).

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

Chetty, Raj, and Emmanuel Saez. 2013. "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients." *American Economic Journal: Applied Economics*, 5.

Congdon, William, Jeffrey Kling, and Sendhil Mullainathan. 2009. "Behavioral Economics and Tax Policy." *NBER Working Paper 15328*.

Farhi, E., and X. Gabaix. 2015. "Optimal Taxation with Behavioral Agents." *Working Paper*.

Feldman, Naomi, and Bradley Ruffle. 2015. "The Impact of Including, Adding and Subtracting a Tax on Demand." *American Economic Journal: Economic Policy*, 2(1): 95–118.

Feldman, Naomi, Jacob Goldin, and Tatiana Homonoff. 2015. "Raising the Stakes: Experimental Evidence on the Robustness of Taxpayer Mistakes." *Working Paper.*

Finkelstein, Amy. 2009. "E-Ztax: Tax Salience and Tax Rates." *The Quarterly Journal of Economics*, 124(3): 969–1010.

Galle, Brian. 2009. "Hidden Taxes." *Washington University Law Review*, 87.

Gamage, David, and Darien Shanske. 2011. "Three Essays on Tax Salience: Market Salience and Political Salience." *Tax Law Review.*

Goldin, Jacob. 2012. "Sales Tax Not Included: Designing Taxes for Inattentive Consumers." *Yale Law Journal.*

Goldin, Jacob, and Tatiana Homonoff. 2013. "Smoke Gets in your Eyes: Cigarette Tax Salience and Regressivity." *American Economic Journal: Economic Policy*, 5(1): 302–336.

Hayashi, Andrew. 2014. "The Legal Salience of Taxation." *University of Chicago Law Review.*

Jones, Damon. 2010. "Withholdings, Salience, and Tax Policy." *Stanford Institute for Economic Policy Research.*

Krishna, Aradhna, and Joel Slemrod. 2003. "Behavioral Public Finance: Tax Design as Price Presentation." *International Tax and Public Finance*, 10(2): 189–203.

Liebman, J., and R. Zeckhauser. 2004. "Schmeduling."

Lipsey, R.G., and Kelvin Lancaster. 1956-67. "The General Theory of the Second Best." *The Review of Economic Studies*, 24.

McCaffery, Edward. 1994. "Cognitive Theory and Tax." *UCLA Law Review*, 41.

O'Donoghue, T., and M. Rabin. 2003. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review Papers and Proceedings*, 93.

Reck, Daniel. 2015. "Taxes and Mistakes: What's in a Sufficient Statistic?" *University of Michigan mimeo.*

Slemrod, Joel, and Wojciech Kopcczuk. 2002. "The Optimal Elasticity of Income." *Journal of Public Economics*, 84.

Appendix

This Appendix derives the formula for incremental adjustments to the balance between high- and low-salience taxes that was discussed in Section 3. Recall that a revenue-neutral shift towards the low-salience tax is defined as a marginal increase in t_l along with whatever change in t_h is required to leave total revenue constant. Define $\theta_i(t_h,t_l) = \frac{\frac{\partial x(p,t_h,t_l)}{\partial t_i}}{\frac{\partial x(p,t_h,t_l)}{\partial p}}$ for $i \in \{h,l\}$, i.e., the ratio of the tax and price derivatives evaluated at taxes t_h and t_l.

Proposition A.1 *Starting at taxes t_h and t_l, a revenue-neutral shift towards the low-salience tax is desirable if and only if $\rho\theta_l(t_h,t_l) + (1-\rho)\theta_h(t_h,t_l) > \theta^*$, where ρ and θ^* are defined as in Proposition 2.*

Proof Suppose that the high and low-salience taxes are set at t_h and t_l. Tracking the derivation of Lemma 2, it is straightforward to show that a revenue neutral shift towards the low-salience tax is welfare improving if and only if

$$U_x(x,y) - p\,U_y(x,y) > 0 \tag{26}$$

The next steps apply a series of Taylor approximations to the quantities in (26).

First, note that

$$U_x\left(x(p,t_h,t_l), y(p,t_h,t_l)\right) \approx U_x^0 + \left(x(p,t_h,t_l) - x(p,0,0)\right)\left(U_{xx}^0 - p\,U_{xy}^0\right) - (t_h+t_l)\,x(p,t_h,t_l)\,U_{xy}^0 \tag{27}$$

where $U_i^0 \equiv U_i^0\left(x(p,0,0), y(p,0,0)\right)$ and $U_{ij}^0 \equiv U_{ij}\left(x(p,0,0), y(p,0,0)\right)$.[20] Similarly,

[20] To reach this result, note that the consumer's budget constraint implies $y(p,t_h,t_l) - y(p,0,0) = -p\left(x(p,t_h,t_l) - x(p,0,0)\right) - (t_h+t_l)\,x(p,t_h,t_l)$. As is common in the literature, this approximation ab-

$$U_y\left(x(p,t_h,t_l),y(p,t_h,t_l)\right) \approx U_y^0+\left(x(p,t_h,t_l)-x(p,0,0)\right)\left(U_{yx}^0-p\,U_{yy}^0\right)-(t_h+t_l)\,x(p,t_h,t_l)\,U_{yy}^0 \tag{28}$$

Next, we can approximate

$$x(p,t_h,t_l) \approx x(p,0,0)+\frac{\partial x}{\partial p}\left(\theta_h t_h+\theta_l t_l\right) \tag{29}$$

where θ_h and θ_l denote $\theta_h(t_h,t_l)$ and $\theta_l(t_h,t_l)$ respectively.

Substituting (27), (28), and (29) into (26) allows us to rewrite (26) as:

$$\frac{\partial x}{\partial p}\left(\theta_h t_h+\theta_l t_l\right)\gamma_0-(t_h+t_l)\,x(t_h,t_l)\,(U_{xy}^0-p\,U_{yy}^0)>0 \tag{30}$$

where $\gamma_0 \equiv U_{xx}^0-2\,p\,U_{yx}^0+p^2\,U_{yy}^0$.

Next, totally differentiating the consumer's budget constraint and first-order condition at the no-tax baseline with respect to I yields $\frac{\partial x}{\partial I}\Big|_{t_h=t_l=0}=-\frac{U_{yx}-pU_{yy}}{\gamma_0}$. Substituting this identity into (30) and rearranging terms allows us to rewrite the condition as $\frac{\partial x}{\partial p}\left(\theta_h t_h+\theta_l t_l\right)+(t_h+t_l)\,\frac{\partial x}{\partial I}<0$. Finally, rewriting in terms of elasticities yields:

$$\varepsilon_{x,p}\,\left(\theta_h t_h+\theta_l t_l\right)>(t_h+t_l)\,\eta_{x,I}\,\omega_x$$

where $\varepsilon_{x,p}$, $\eta_{x,I}$, and ω_x are defined as in Proposition 2. Dividing both sides of the equation by $\varepsilon_{x,p}\,(t_h+t_l)$ and applying the definitions of ρ and θ^* yields the result. ■

stracts from third-order and higher terms.

Figure 1: Illustration of Result