Dynamic Corrective Taxes with Time-Varying Salience[☆]

Ben Gilbert[a,*], Joshua S. Graff Zivin[b]

[a] *Colorado School of Mines, 1500 Illinois St, Golden CO, 80401, United States*
[b] *University of California, San Diego and NBER, 9500 Gilman Dr, La Jolla CA, 92093-0519*

Abstract

The intermittency of payment for many goods creates a disconnect between paying and consuming such that the marginal price is not always salient when consumption decisions are made. This paper derives optimal dynamic corrective taxes when there are externalities as well as internalities from inattention and persistence in consumption across periods. Our optimal taxes address dynamic inefficiencies that are not captured in static models of inattention. We also characterize a second-best constant tax and the excess burden associated with time-invariant tax rates. We then calibrate the model to U.S. residential electricity consumption.

JEL Codes: D03, D11, D62, D91, H21, H23, L97, Q40, Q41, Q50

Keywords: Salience; Inattention; Optimal Taxes; Energy Demand; Consumption Persistence

[☆] We thank Hunt Allcott, Terry Iverson, Jacob LaRiviere, Andrea La Nauze, Pete Maniloff, Chuck Mason, Kevin Novan, Linda Thunstrom, Klaas van 't Veld, and numerous seminar participants for helpful comments. Any errors are our own.

[*] Corresponding author. ph: 1 303 384 2359, fax: 1 303 273 3416

Email addresses: bgilbert@mines.edu (Ben Gilbert), jgraffzivin@ucsd.edu (Joshua S. Graff Zivin)

1. Introduction

The standard economic remedy to externality problems has been well known since the 1920s and has been a regular fixture in the policy makers toolkit since at least the 1980s. The idea is simple, yet elegant. Impose a tax equal to marginal external damages in order to internalize externalities and generate private decision making that is socially optimal (Pigou, 1920). But in an increasingly cashless society where the financial consequences of ones daily choices may only be experienced monthly or even quarterly, is it realistic to expect consumers to optimally perform this calculation? Indeed, an emerging literature that examines the impacts of price salience on purely private decisions suggests that this is unlikely to be the case across a range of contexts (DellaVigna, 2009; Finkelstein, 2009; Sexton, 2015; Grubb and Osborne, 2015; Karlan, et. al., 2016).

This so-called inattention problem raises the specter of additional policy interventions that increase salience in order to fix the internality from privately suboptimal decision making (Chetty et al., 2009; Chetty, 2009; Allcott et al., 2014; Allcott and Taubinsky, 2015; Farhi and Gabaix, 2018).[1] The implications for a simple static setting are straightforward – impose a super tax that forces consumers to face the full internal (i.e. salient-equivalent) and external costs. This logic becomes more complex in a dynamic setting with a more realistic representation of price salience as intermittent. The optimal (time-varying) tax will now depend on both the persistence of consumption across periods and the degree to which consumers are forward-looking in their behavior. Consumption persistence can arise for a host of reasons, including habits, status quo bias, or complementarity with durable goods (e.g. Becker and Murphy, 1988; Benhabib and Bisin, 2005; Flavin and Nakagawa, 2008; Landry, 2018). Moreover, recent research has shown that consumption persistence is an important consideration in the welfare effects of a corrective policy (Costa and Gerard, 2018). Our work extends their key insights by showing that these welfare effects are compounded by time-varying inattention and demonstrating the importance of both persistence and inattention for the design of optimal and second-best policy. Importantly for policy design, this persistence provides a vehicle through which forward-looking agents can commit to future paths of consumption.

In this paper, we develop a model of consumer behavior when prices are intermittently salient, demand persists across periods, and consumption creates external damages. We

[1] This internality is conceptually similar to the one that arises in the time-inconsistency models of taxation. See Gruber and Koszegi (2004) and O'Donoghue and Rabin (1999, 2006) for example.

then derive optimal dynamic tax rules for three distinct types of consumers: a *fully naive* agent who is myopic about her consumption persistence and not sophisticated about future inattention; a *partially naive* agent who dynamically optimizes her consumption persistence under perceived prices but is not sophisticated about future price inattention; and a *sophisticated* agent who dynamically optimizes her consumption persistence under perceived prices, with sophistication about future inattention to prices. Our work builds on the static models of corrective policy with externalities and internalities developed by Allcott, et al. (2014) and Allcott and Taubinsky (2015) by evaluating a fully dynamic setting with time-varying inattention and consumption persistence.[2]

We find that the optimal time-varying tax depends upon the price-elasticity of demand and the size of external damages, as well as a number of parameters that govern the dynamic nature of the problem including the attention decay function, price salience, consumption persistence, and time preference parameters. In general, sophisticated types face lower taxes than either of the naive types, with the size of these differences quite sensitive to the habit persistence and salience parameters. Greater intertemporal consumption linkages resulting from habit persistence lead to higher taxes for the fully naive agent during all periods; lower taxes for the sophisticated one during the salient period only; and have no effect on taxes for the partially naive individual. The effects of salience are slightly different. Greater levels of inattention lead to higher taxes for all agents during non-salient periods. During salient periods, greater levels of future inattention have no effect on the optimal tax for fully or partially naive agents but lead to lower taxes for sophisticated ones.

When regulators are constrained to a time-invariant tax, the optimal constant tax is second-best. We find that the second-best constant tax rate is an exact weighted-average of the optimal dynamic tax in each period, in which the weights now depend on the parameters for price-elasticity of demand, the attention decay function, price salience, consumption persistence, and time preference. This second-best tax balances under-consumption during the salient period with over-consumption during the non-salient period. Second-best taxes are lowest for sophisticated agents and highest for fully naive agents.

Because time-invariant taxes are suboptimal, the welfare effects of suboptimal consumption in each period spill over into each other time period through realized or anticipated

[2]Allcott, et al. (2014) and Allcott and Taubinsky (2015) model adoption of a durable good whose utilization may generate externalities. Although durable goods have dynamic implications in the sense that the impacts of decisions will be felt across time, adoption and utilization are one-shot decisions in these models. In contrast, we model a recurrent purchase over an infinite time horizon with periodically salient price signals.

consumption persistence. We derive formulas for these welfare effects that incorporate dynamic spillovers and can be calculated for any time-invariant tax rate. We then evaluate these welfare formulas under three candidate tax rates: our second-best constant tax, a "static" tax that is the optimal tax from a static model of price salience and therefore ignores dynamic inefficiency, and the standard Pigouvian tax equal to the marginal externality. The welfare losses from time-invariant taxes are largest for fully naive agents.

Our theoretical work is following by a numerical calibration for U.S. residential electricity consumption. The optimal tax for the fully naive agent is several times larger than that for the sophisticated one. The optimal tax for the partially naive agents lies between them. In all cases, the optimal tax rises as attention decays, arriving at a level between three and twelve times larger than the simple Pigouvian tax that neglects inattention. The second-best tax regime, which leads to lower-than-optimal tax rates during non-salient periods and higher-than-optimal tax rates during salient periods, exacerbates the differences across agent types. The excess burden from the second-best tax ranges from about 19 dollars per household for partially naive agents to about 115 dollars per household for fully naive agents. With approximately 126 million households in the U.S., this amounts to between 2.4 billion and 14.5 billion dollars in deadweight loss from the electricity sector, even if the second-best optimal tax is implemented. These welfare losses rise sharply if the tax is reduced from the second-best level, if inattention is more severe, or if demand is more elastic. For example, the welfare loss associated with having no tax at all ranges from 181 dollars per household for sophisticated agents to 4,560 dollars per household for fully naive agents.

The key insights from our model are not limited to electricity markets and should be relevant for any price-based policy designed to address market failures, where price/cost inattention and linked intertemporal decision making are commonplace. For example, gasoline consumption 1) emits several types of pollutants; 2) depends on vehicle type and driving habits that persist across many periods; and 3) does not present drivers with explicit, salient prices when making trip-level decisions. Similarly, unhealthy food choices 1) impose health costs on society; 2) are sometimes habit-forming, such that enjoyment in the current period depends on consumption in prior periods; and 3) are generally made at higher frequency than the purchases that make their costs salient. The increasing prevalence of auto-billing and subscription services for a range of purchases will only increase its applicability.

In the next section, we set up the model, derive optimal and second-best taxes, and derive formulas for the welfare loss from time-invariant taxes. The subsequent section presents the results from the numerical simulation. The final section discusses and concludes.

2. Model

We first describe the economic environment and characterize the private solution for three types of agents:

- a *fully naive* agent who is myopic about her habit persistence and not sophisticated about future inattention;
- a *partially naive* agent who dynamically optimizes her habit persistence under perceived prices but is not sophisticated about future price inattention;
- a *sophisticated* agent who dynamically optimizes her habit persistence under perceived prices, with sophistication about future inattention to prices.

2.1. Model Setup

Utility in each period depends on consumption of a clean numeraire good y_t, a dirty good x_t, and past consumption of the dirty good, αx_{t-1}. The parameter α governs the persistence of consumption decisions across time. Consumption of the dirty good produces social damages proportional to current consumption that last only for the current period:

$$D_t = \phi x_t. \tag{1}$$

Here, ϕ is the marginal external cost of damages associated with consumption of x.

For simplicity we model period utility as quasilinear:

$$U_t = u_t(x_t - \alpha x_{t-1}) + y_t \tag{2}$$

where subscripts on U_t and u_t are used only to index the time period (the functional forms are assumed to be the same in all periods). Consumption of x in adjacent time periods are "adjacent complements" in the sense that as x_{t-1} increases, the marginal utility of x_t rises in period t and consumption of x_t increases *ceteris paribus*. This type of intertemporal dependence can arise if a good is habit-forming (Becker and Murphy, 1988), when consumption is based on defaults or status quo bias (Samuelson and Zeckhauser, 1988) or when consumption depends on some durable good stock (Flavin and Nakagawa, 2008). For simplicity, we will refer to the dirty good as habit forming, where the α parameter governs consumption persistence through the effect of past consumption on the utility of current consumption. This provides the agent a commitment mechanism by which to mitigate potential future overconsumption by reducing present consumption in order to change habits, if the agent is sophisticated about future inattention[3]. We ignore saving and borrowing so that the household faces a budget constraint in each period:

$$m = px_t + y_t, \tag{3}$$

where m denotes per period household wealth and p is the price of the dirty good. This simplification allows us to focus on the dynamic relationship from hysteresis rather than the intertemporal transfer of financial resources[4].

The price of the dirty good is intermittently salient. To fix ideas, consider a monthly bill. Agents receive a salient price signal in period 1, which we can think of as receiving a bill on the first day of the billing cycle. The price may remain salient for a number of periods $i = 1, ..., I$ with $I < T$. Let us call this the "salient window". The price is definitively not salient from period $I + 1$ until the end of the cycle in period T - the "insalient window". In period $T+1$ the agents receive another salient price signal (a new bill) and the cycle repeats. Agents live for an infinite number of cycles indexed by M.

It is our contention that agents are more attentive to prices the more recently the price signal was received. We model this feature of attention decay by assuming that the probability of perceiving the true price is equal to 1 in period 1 and declines to zero by period I as follows:

$$\mu_i = 1 - \frac{i-1}{I-1}$$

With probability μ_i, the agent remembers the price signal in period i and optimizes with respect to the true price. With probability $(1-\mu_i)$, the agent is not thinking about the price signal in period i - the price is not salient to the agent - and the agent makes optimization errors.

If the price of x is not salient, agents are less responsive to prices than when they are fully salient, and therefore overconsume the dirty good.[5] We represent this by having agents

[3]This model has a similar structure to a model of status quo bias with an endogneous status quo, in which past choices affect the current default decision. Introducing borrowing and savings, a single permanent income constraint, or durable goods stock accumulation, would provide a similar dynamic mechanism. We focus on habits because they are an empirical feature of the polluting goods we are interested in, and because intermittent price salience has important implications for habit formation.

[4]This simplification also allows us to avoid theoretical cases where a sophisticated household could be so concerned about future overconsumption that it would "raid" its savings in the present and leave nothing to be wasted by suboptimal decisions in the future.

[5]Although it is possible that agents underconsume when prices are not salient, empirical evidence suggests

undervalue the true price by a salience factor $\theta \in (0,1)$ and optimize as if their period budget constraint was:

$$m = \theta p x_t + y_t \tag{4}$$

During the salient window, the agent correctly perceives the budget constraint in equation (3) with probability μ_i, and with probability $(1-\mu_i)$ incorrectly acts as if equation (4) is the budget constraint. During the insalient window the perceived budget constraint is equation (4). This cycle repeats for periods $T+1$ to $2T$, and so on.

2.2. Private optimization

In the first time period, when the perceived price corresponds to the true price, the agent optimizes the following equation:

$$\begin{gathered} \max_{\{x_{tM}, y_{tM}\}} \sum_{M=1}^{\infty} \sum_{t=1}^{T} \beta^{T(M-1)+t-1} \left[U_{tM}(x_{tM}, y_{tM}, x_{t-1,M}) - \lambda_{tM}(p x_{tM} + y_{tM} - m)\right] \\ s.t.\ x_0 \text{ given} \end{gathered} \tag{5}$$

where β is the discount factor and M indexes a cycle of length T periods, i.e., a billing cycle.

In contrast, once I periods have passed, the price is definitively not salient and the agent's maximization problem can be expressed as:

$$\begin{aligned} \max_{\{x_t, y_t, x_{sM}, y_{sM}\}} \quad & \sum_{t=I+1}^{T} \beta^{t-I-1} \left[U_t(x_t, y_t, x_{t-1}) - \lambda_t(\theta p x_t + y_t - m)\right] \\ & + \sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[U_{sM}(x_{sM}, y_{sM}, x_{s-1,M}) - \lambda_{sM}(\theta p x_{sM} + y_{sM} - m)\right] \\ & s.t.\ x_0, ..., x_I \text{ given} \end{aligned} \tag{6}$$

From the point of view of period $I+1$, the agent perceives the price to be θp and applies this price to their optimal consumption plan through the end of the current billing cycle and all future cycles $M = 2, ..., \infty$.

During periods $1 < i \leq I$ (the salient window), with probability μ_i the agent perceives the true price and solves a problem of the type in equation (5); with probability $(1-\mu_i)$ the agent is inattentive to the price and solves a problem of the type in equation (6).

When setting optimal taxes, the regulator does not know whether the agent is in an attentive or inattentive state in period i. The regulator is therefore interested in the solution to the expected objective function

---

overconsumption is generally prevalent.

$$
\begin{aligned}
\max_{\{x_t, y_t, x_{sM}, y_{sM}\}} \quad & \mu_i \Big\{ \textstyle\sum_{t=i}^{T} \beta^{t-i} \left[U_t(x_t, y_t, x_{t-1}) - \lambda_t (p x_t + y_t - m) \right] \\
& + \textstyle\sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[U_{sM}(x_{sM}, y_{sM}, x_{s-1,M}) - \lambda_{sM}(p x_{sM} + y_{sM} - m) \right] \Big\} \\
& + (1-\mu_i) \Big\{ \textstyle\sum_{t=i}^{T} \beta^{t-i} \left[U_t(x_t, y_t, x_{t-1}) - \lambda_t (\theta p x_t + y_t - m) \right] \\
& + \textstyle\sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[U_{sM}(x_{sM}, y_{sM}, x_{s-1,M}) - \lambda_{sM}(\theta p x_{sM} + y_{sM} - m) \right] \Big\} \\
& s.t. \; x_0, ..., x_{i-1} \text{ given}
\end{aligned} \tag{7}
$$

As time passes following the initial salient price signal, the agent becomes increasingly likely to optimize relative to the perceived price θp rather than the true price p. The regulator incorporates this transition into the optimal tax sequence as we will show.

2.2.1. Fully Naive and Partially Naive Agents

The fully naive agent takes x_{t-1} as given in *each* time period and is not forward-looking about the persistent impacts of today's consumption on tomorrow's utility. Instead, the agent passively adapts to historical consumption and optimizes under current perceived prices.[6] The consumption plan chosen during the insalient window ($I+1 \leq t \leq T$) by the fully naive household is defined by the first order conditions:

$$
\begin{aligned}
u'_t &= \theta p, \quad t = I+1, ..., T \\
u'_{sM} &= \theta p, \quad s = 1, ..., T; \;\; M = 2, ..., \infty
\end{aligned} \tag{8}
$$

We model partially naive agents as Becker-Murphy rational habit formers who are naive about future inattention. We assume they are forward-looking about consumption persistence. This feature is consistent with empirical evidence of consumption persistence, including estimates of habit persistence and cases in which consumption persistence occurs through investment in a complementary durable good. In this case agents consider how today's decisions affect their future demands, even if they fail to recognize how their price inattention will change in the future. The first order conditions for the consumption plan made by the partially naive agent during the insalient window are

[6]An alternative example to habit persistence is status quo bias. If status quo behavior is a function of decision history, then decisions will be persistent but will not respond to expected future changes in the status quo.

$$
\begin{aligned}
u'_t - \alpha\beta u'_{t+1} = \quad & \theta p, \qquad t = I+1, ..., T \\
u'_{sM} - \alpha\beta u'_{s+1,M} = \quad & \theta p, \qquad s = 1, ..., T; \;\; M = 2, ..., \infty
\end{aligned}
$$

During the salient window, $t = i \leq I$, the probability of perceiving the true price is μ_i. The regulator's objective is to set an optimal period i tax that incorporates perceived prices. Neither of the naive agent types foresees insalient prices ($\theta < 1$) in future periods, and the fully naive agent also does not foresee the effect of today's consumption on future demand. From the point of view of the regulator, we can represent the period i objective function for either of these agent types as

$$
\begin{aligned}
\underset{\{x_t, y_t, x_{sM}, y_{sM}\}}{max} \quad & \mu_i \Big\{ \textstyle\sum_{t=i}^{T} \beta^{t-i} \left[u_t(x_t - \alpha\overline{x}_{t-1}) + y_t - \lambda_t(px_t + y_t - m) \right] \\
& + \textstyle\sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[u_{sM}(x_{sM} - \alpha\overline{x}_{s-1,M}) + y_t - \lambda_{sM}(px_{sM} + y_{sM} - m) \right] \Big\} \\
& + (1-\mu_i) \Big\{ \textstyle\sum_{t=i}^{T} \beta^{t-i} \left[u_t(x_t - \alpha\overline{x}_{t-1}) + y_t - \lambda_t(\theta px_t + y_t - m) \right] \\
& + \textstyle\sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[u_{sM}(x_{sM} - \alpha\overline{x}_{s-1,M}) + y_t - \lambda_{sM}(\theta px_{sM} + y_{sM} - m) \right] \Big\} \\
& s.t. \; x_0, ..., x_{i-1} \text{ given}
\end{aligned} \tag{9}
$$

The fully naive agent is myopic about consumption persistence and takes $\overline{x}_{t-1}$ as given in each period when optimizing (9). Partially naive agents, on the other hand, dynamically optimize (9) over $\overline{x}_{t-1} = x_{t-1}$.

In the fully naive case, during the salient window the regulator faces an agent whose consumption plan solves

$$
\begin{aligned}
u'_t = \quad & p, \quad t = i, ..., T \\
u'_{sM} = \quad & p, \quad s = 1, ..., T; \;\; M = 2, ..., \infty
\end{aligned} \tag{10}
$$

with probability μ_i, and with probability $(1-\mu_i)$ the agent's consumption plan solves

$$
\begin{aligned}
u'_t = \quad & \theta p, \quad t = i, ..., T \\
u'_{sM} = \quad & \theta p, \quad s = 1, ..., T; \;\; M = 2, ..., \infty.
\end{aligned} \tag{11}
$$

In effect, the first order condition that is relevant for the regulator's tax-setting problem is

$$
\begin{aligned}
u'_t = \quad & p \cdot (\mu_i + (1-\mu_i)\theta), \quad t = i, ..., T \\
u'_{sM} = \quad & p \cdot (\mu_i + (1-\mu_i)\theta), \quad s = 1, ..., T; \;\; M = 2, ..., \infty
\end{aligned} \tag{12}
$$

Note that if either $\mu_i = 1$ or $\theta = 1$, the marginal utility is equal to the price as in the standard model.

For partially naive agents during periods during the salient window, the first order condition relevant for the regulator's tax-setting problem is

$$\begin{aligned} u'_t - \alpha\beta u'_{t+1} = &\ \ p \cdot (\mu_i + (1-\mu_i)\theta), \quad t = i, ..., T \\ u'_{sM} - \alpha\beta u'_{s+1,M} = &\ \ p \cdot (\mu_i + (1-\mu_i)\theta), \quad s = 1, ..., T; \ \ M = 2, ..., \infty \end{aligned} \tag{13}$$

where, in comparison to the fully naive agent, the partially naive agent considers the effect of consumption persistence on future utility.

2.2.2. Sophisticated Agents

From the standpoint of a period in the insalient window, the sophisticated agent's solution to (6) is forward-looking about habit formation but is subject to optimization errors from inattention (θ):

$$\begin{aligned} u'_t - \alpha\beta u'_{t+1} = &\ \ \theta p, \quad t = I+1, ..., T \\ u'_{sM} - \alpha\beta u'_{s+1,M} = &\ \ \theta p, \quad s = 1, ..., T; \ \ M = 2, ..., \infty \end{aligned} \tag{14}$$

These first order conditions are identical to those of the partially naive agent during the insalient window; the agent behaves in a sophisticated way only during periods in which they are attentive to the true price.

By recursive substitution of (14), we obtain

$$\begin{aligned} & u'_{tM} = (\alpha\beta)^{\tau-1} + \theta p \textstyle\sum_{s=0}^{\tau}(\alpha\beta)^s \\ \Longrightarrow\ & u'_{tM} \approx \theta p \textstyle\sum_{s=0}^{\infty}(\alpha\beta)^s = \theta p R \end{aligned} \tag{15}$$

where $R = \sum_{s=0}^{\infty}(\alpha\beta)^s = \frac{1}{1-\alpha\beta}$.

Let the consumption plan defined by (15) be

$$\hat{x}_{tM} = \hat{x}_{tM}(x_{t-1,M}, p; \theta), \quad \hat{y}_{tM} = \hat{y}_{tM}(x_{t-1,M}, p; \theta), \tag{16}$$

where this is the consumption *plan* made in period $I+1$ of the current cycle that the agent expects to carry out from $I+1$ through all future periods and all future cycles. Note that $\hat{x}_{tM}$ is larger than the privately optimal demand with fully salient prices (and $\hat{y}_{tM}$ is smaller) because x is overconsumed out of the budget when the price is not salient.

Consider the agent's decision during the salient window, when agents may or may not perceive the true price. We model sophistication about inattention as follows. If the agent does not perceive the true price in some period $t \leq I$ during the salient window, the agent will behave according to (15). If the agent perceives the true price in the current period, then the agent expects to *pay* that price for consumption in the current and all future periods, but expects to make future *consumption decisions* according to (16). In addition, she considers how the current consumption decision will shift those future demand functions. These future demands enter the agent's effective period i objective function as follows:

$$\max_{\{x_i,y_i\}} \mu_i \Bigg\{ U_i(x_i, y_i; x_{i-1}) - \lambda_i(px_i + y_i - m)$$
$$+ \sum_{t=i+1}^{T} \beta^{t-i-1} \left[U_t(\hat{x}_t(x_i), \hat{y}_t(x_i), \hat{x}_{t-1}(x_i)) - \lambda_t(p\hat{x}_t(x_i) + \hat{y}_t(x_i) - m)\right]$$
$$+ \sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[U_{sM}(\hat{x}_{sM}(x_i), \hat{y}_{sM}(x_i), \hat{x}_{s-1,M}(x_i)) - \lambda_{sM}(p\hat{x}_{sM}(x_i) + \hat{y}_{sM}(x_i) - m)\right] \Bigg\}$$
$$+ (1-\mu_i) \Bigg\{ U_i(x_i, y_i; x_{i-1}) - \lambda_i(\theta px_i + y_i - m)$$
$$+ \sum_{t=i+1}^{T} \beta^{t-i-1} \left[U_t(\hat{x}_t(x_i), \hat{y}_t(x_i), \hat{x}_{t-1}(x_i)) - \lambda_t(\theta p\hat{x}_t(x_i) + \hat{y}_t(x_i) - m)\right]$$
$$+ \sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \left[U_{sM}(\hat{x}_{sM}(x_i), \hat{y}_{sM}(x_i), \hat{x}_{s-1,M}(x_i)) - \lambda_{sM}(\theta p\hat{x}_{sM}(x_i) + \hat{y}_{sM}(x_i) - m)\right] \Bigg\}$$
$$s.t. \quad x_0, ..., x_{i-1} \text{ given} \qquad (17)$$

The first order condition for x_i is

$$\left[u'_i - \alpha\beta u'_{i+1}\right] + \mu_i \Bigg(\sum_{t=i+1}^{T} \beta^{t-i-1} \frac{\partial \hat{x}_t}{\partial x_i} \left[u'_t - \alpha\beta u'_{t+1} - p\right]$$
$$+ \sum_{M=2}^{\infty} \sum_{s=1}^{T} \beta^{T(M-1)+s-1} \frac{\partial \hat{x}_{sM}}{\partial x_i} \left[u'_{sM} - \alpha\beta u'_{s+1,M} - p\right] \Bigg)$$
$$= p \cdot (\mu_i + (1-\mu_i)\theta) \qquad (18)$$

The first term in brackets is the marginal utility across time taking consumption persistence into account. This would be equal to the price during period 1 (when prices are fully salient) if the agent did not demand commitment to avoid future inattention, captured by the effects of x_i on future $\hat{x}$. Each term in brackets inside the large parentheses is negative. These terms measure the extent of the deviation from a private optimum in future periods due to salience effects. Because the agent is overconsuming in the future, the marginal utilities across time fall below the price. The sophisticated agent anticipates these deviations. Denote the period i demand that solves equation (18) as $\hat{x}_i(x_{i-1}, p; \theta)$. This is lower than the privately optimal demand would be in the absence of salience effects, as the sophisticated agent tries to use persistence as a commitment device for a lower consumption path.

By substituting the conditions in (14) into (18) and simplifying, we obtain

$$\begin{aligned} u_i' - \alpha\beta u_{i+1}' &= p\mu_i(1-\theta)\left(\sum_{t=i+1}^{T} \beta^{t-i-1}\frac{\partial \hat{x}_t}{\partial x_i} + \sum_{M=2}^{\infty}\sum_{s=1}^{T} \beta^{T(M-1)+s-1}\frac{\partial \hat{x}_{sM}}{\partial x_i} \right) + p\cdot(\mu_i + (1-\mu_i)\theta) \\ &= p\left(\mu_i(1+\delta) + (1-\mu_i)\theta\right) \end{aligned} \tag{19}$$

where $\delta = (1-\theta)\left(\sum_{t=i+1}^{T} \beta^{t-i-1}\frac{\partial \hat{x}_t}{\partial x_i} + \sum_{M=2}^{\infty}\sum_{s=1}^{T} \beta^{T(M-1)+s-1}\frac{\partial \hat{x}_{sM}}{\partial x_i} \right)$ is the discounted impact of today's consumption decision on all future demand functions. With quasilinear utility, terms $\frac{\partial \hat{x}_t}{\partial x_i}$ are constant demand shifters, so δ is a constant. If there are no future salience effects to avoid ($\theta = 1$), or if there are no habits (or no commitment mechanism by which to avoid salience effects, i.e., $\alpha = 0$) then $\delta = 0$. If in addition $\mu_i = 1$, then the problem collapses to a standard period 1 private optimum for habit formation with the price on the right hand side.

2.2.3. Summary of privately optimal consumption paths

In the absence of internalities from inattention and naivety about persistence, equating the left hand side of (19) to the price in all time periods will yield the true privately optimal consumption path. By contrast, sophisticated agents underconsume relative to the private optimum when μ_i is close to one and overconsume as μ_i approaches zero, so that consumption rises in expectation throughout the salient window. Comparing the sophisticated agent conditions in (19) to those for the partially naive agent in (13), notice that they differ by the presence of $(1+\delta)$. Although consumption also rises throughout the salient window for partially naive agents, these agents consume at the true private optimum in the first period and then overconsume in all other periods. Lastly, comparing the sophisticated agent's conditions for periods during the insalient window in (15) to those for the fully naive agent in (8), we can see that they differ by the presence of R for the sophisticated agents. The marginal utility of consumption is therefore larger for sophisticated agents than fully naive agents during the insalient window, which indicates that overconsumption in the later part of the billing cycle is more severe for fully naive agents. Similarly during the salient window the marginal utility of consumption differs in each period by a factor of R; for the fully naive agent the marginal utility is equated to $p \cdot (\mu_i + (1-\mu_i)\theta)$ in expectation whereas the sophisticated agent sets marginal utility equal to $p \cdot (\mu_i + (1-\mu_i)\theta) \cdot R$ in expectation. Again, the fully naive agent always consumes more than the sophisticated agent. Further, the true private optimum sets marginal utility in a given period equal to $p + \alpha\beta u'_{t+1}$. Therefore, the fully naive agent also always overconsumes relative to the true private optimum.

2.3. Optimal Dynamic Corrective Taxes

Consumption will deviate from the socially optimal path because of the internalities outlined at the end of the previous subsection, and because of externalities (ϕ) generated by consumption of the dirty good. A system of optimal corrective taxes must address both sources of inefficiency. In this subsection we set up the social planner's problem and solve for the dynamic corrective taxes that induce each agent type to consume along the socially optimal path.

Assuming a rate of time preference equal to the interest rate, the social planner chooses a path $\{x_t, y_t\}_{t=1}^{\infty}$ to maximize the intertemporal social welfare function

$$W = \sum_{t=1}^{\infty} \beta^{t-1} \left[U_t(x_t, y_t, x_{t-1}) - \phi x_t - \lambda_t (p x_t + y_t - m) \right] \tag{20}$$

subject to x_0 given. Note that the social planner optimizes over all time periods regardless of the billing cycle. For the rest of this subsection we drop the M subscript and use the t and i subscripts to denote the position within a given cycle. The optimal path is defined by the first order conditions:

$$u'_t - \alpha\beta u'_{t+1} = p + \phi, \quad t = 1 ... \infty \tag{21}$$

One implication of (21) is that absent inattention or naivety about persistence, a tax τ at the standard Pigouvian rate of $\tau = \phi$ would induce households to consume at the social optimum. By recursive substitution of (21), we obtain

$$\begin{aligned} u'_t &= (\alpha\beta)^{\tau-1} u'_\tau + (p+\phi) \textstyle\sum_{s=0}^{\tau} (\alpha\beta)^s \\ &\approx (p+\phi) \textstyle\sum_{s=0}^{\infty} (\alpha\beta)^s = (p+\phi) R. \end{aligned} \tag{22}$$

The socially optimal per-period marginal utility is increased (and per-period consumption is restrained) by consideration of both the external costs ϕ and the recursive impact of today's consumption on future consumption through habit persistence, or the effect of α through R.

Let $j \in \{F, P, S\}$ denote fully naive, partially naive, and sophisticated agent types, respectively. Each agent type j could be induced to consume along the optimal path by a dynamic tax that transforms their private first order conditions in each period into the social planner's first order conditions from (22).

During the insalient window of the billing cycle, this would be an extra large tax $\tau_{tj}(\theta)$ which is different for agent type j. For fully naive agents F the optimal tax must solve

$$u'_t = \theta(p + \tau_{tF}(\theta)) = (p + \phi)R, \quad t = I + 1, ..., T, \tag{23}$$

whereas for partially naive and sophisticated agents the optimal tax must solve

$$u'_{tM} = \theta(p + \tau_{tj}(\theta))R = (p + \phi)R, \quad j \in \{P, S\}, \ t = I + 1, ..., T. \tag{24}$$

The optimal taxes during the salient window of the billing cycle can be found in a similar manner. For fully naive agents, the optimal tax must solve

$$u'_i = (p + \tau_{iF}(\theta)) \cdot (\mu_i + (1 - \mu_i)\theta) = (p + \phi)R, \quad i = 1, ..., I, \tag{25}$$

for partially naive agents, the optimal tax solves

$$u'_i = (p + \tau_{iP}(\theta)) \cdot (\mu_i + (1 - \mu_i)\theta)\, R = (p + \phi)R, \quad i = 1, ..., I, \tag{26}$$

and for sophisticated agents the optimal tax solves

$$u'_i = (p + \tau_{iS}(\theta)) \cdot (\mu_i(1 + \delta) + (1 - \mu_i)\theta)\, R = (p + \phi)R, \quad i = 1, ..., I, \tag{27}$$

The set of taxes that solves all of these conditions is exactly that which transforms each agent's first order conditions from equations (8), (9), and (15) in periods $t = I + 1, ..., T$, and from equations (12), (13), and (19) in periods $i = 1, ..., I$, into the social planner's first order conditions in (22). The solution is stated in the following proposition.

Proposition 1. *The optimal corrective taxes are time-varying and their magnitudes depend on the forward-looking behavior of the households, i.e., whether households are fully naive (F), partially naive (P), or sophisticated (S), as well as how recently a salient price signal was received. In order to conserve notation and display a single set of formulas for all agent types, let* $R_j = R = \frac{1}{1-\alpha\beta}$ *for fully naive agents and* $R_j = 1$ *for sophisticated and* *partially naive agents. Further, let $\delta_j = 0$ for fully and partially naive agents and $\delta_j = \delta$ for sophisticated agents. Then for $t = I+1, ..., T$,*

$$\tau_{tj}(\theta) = p\frac{R_j - \theta}{\theta} + \phi\frac{R_j}{\theta} > \phi, \tag{28}$$

and for $i = 1, ..., I$,

$$\tau_{ij}(\theta) = p\frac{R_j - (\mu_i(1+\delta_j) + (1-\mu_i)\theta)}{\mu_i(1+\delta_j) + (1-\mu_i)\theta} + \phi\frac{R_j}{\mu_i(1+\delta_j) + (1-\mu_i)\theta} <> \phi. \tag{29}$$

Several comments about the optimal dynamic corrective taxes are in order. First, the optimal tax in each period can generally be considered a weighted combination of exterality correction and price salience internality correction, with weights depending on how forward-looking the agent is through (R and δ). This can be seen most clearly during the first period (or if otherwise either $\mu_i = 1$ or $\theta = 1$) when the taxes are

$$\begin{aligned}
\tau_{1F}(\theta) = &\quad p(R-1) + \phi R > \phi \\
\tau_{1P}(\theta) = &\quad \phi \\
\tau_{1S}(\theta) = &\quad p\tfrac{-\delta_1}{1+\delta_1} + \phi\tfrac{1}{1+\delta_1} < \phi.
\end{aligned}$$

Even when prices are salient in the first period, the fully naive tax is larger than marginal external damage because of the internality of passive habit persistence; the partially naive agent does not experience this internality, so their tax is equal to marginal external damage when prices are salient; and the sophisticated agents try to avoid future internalities from inattention by reducing consumption when prices are salient, so $\tau_{1S}(\theta)$ (and $\tau_{iS}(\theta)$ for period i soon enough after period 1) is less than the marginal social damage and can even be a subsidy.

Second, the importance of the inattention parameter θ varies across time and agent type. For fully and partially naive agents, the taxes in the initial fully salient period do not depend on θ; at $t = 1$ naive agents *do not plan* to be inattentive. Therefore, they do not demand commitment, and the optimal tax cannot influence their demand for such commitment. A consequence of this for fully naive agents is that if there is no habit persistence and $R = 1$, then the first period tax is equal to the marginal externality, $\tau_{1F}(\theta) = \phi$. If there are habits then $R > 1$ and the first period tax additionally corrects the internality of failing to rationally plan for habits ($p(R-1)$) as well as the additional externalities caused by the internality (ϕR).

A smaller θ implies a larger tax for all agent types during later periods in which agents are not fully attentive to the true price. This occurs because the perceived price becomes farther from the true price. During earlier periods when the true price is more likely to be salient, however, the relationship between θ and the optimal tax depends on how forward-looking the agent is. A smaller θ requires an unambiguously larger tax for fully and partially naive agents during these earlier periods, but for sophisticated agents a smaller θ also increases δ, which captures the anticipated effect of today's consumption on all future demand functions. Sophisticated agents anticipate greater overconsumption the smaller is θ, further reducing consumption in earlier periods, which requires a smaller tax during the earlier periods when perfect price attention is more likely.

Third, the optimal tax for all agent types increases over the salient window and the relative ranking across agent types is preserved throughout the billing cycle. The optimal tax increases because the probability of perceiving the true price declines, with the largest tax occuring in the insalient window. The relative ranking is preserved at every time step because the overconsumption from internalities is always the greatest for fully naive agents, and always at least as great for partially naive than sophisticated agents; partially naive agents only overconsume during inattentive periods, but never anticipate this overconsumption. Further, the internalities exacerbate the externality because they lead to excessive consumption of a dirty good. The optimal corrective tax on the sophisticated agent is therefore never larger than the tax for the partially naive agent, and the tax for the fully naive agent is strictly greater in each period than the tax for the other two agents:

$$\begin{aligned}
\tau_{tF}(\theta) > \quad & \tau_{tP}(\theta) = \tau_{tS}(\theta) && > \phi, \quad t = I+1, ..., T, \\
\tau_{iF}(\theta) > \quad & \tau_{iP}(\theta) && > \tau_{iS}(\theta), \quad i = 2, ..., I, \\
\tau_{1F}(\theta) > \quad & \tau_{1P} = \phi && > \tau_{1S}(\theta)
\end{aligned} \tag{30}$$

Fourth, the relationship between habit persistence (α) and the optimal tax differs across agent types and across time within the billing cycle. For the fully naive agent, if there is no habit persistence then $\alpha = 0$ and $R = 1$, so that the fully naive agent's tax $\tau_{tF}(\theta)$ does not need to address the internality from being passive about consumption persistence. If instead $\alpha > 0$ (and $R > 1$), the optimal tax for the fully naive agent $\tau_{tF}(\theta)$ must correct both inattention to prices and myopia about consumption persistence in order to induce the optimal path.

For the partially naive agents, habit persistence is not passive so the tax only needs to correct the inattention internality and the pollution externality. We can see this because R -

the cumulative impact of persistence - is not present in the partially naive tax in any period; the partially naive agent has already optimized over habit persistence and no tax correction is required. The regulator can therefore put the partially naive agent on the optimal path when the salient price signal is received ($t = 1$) by just correcting the externality, then keep her on the optimal path by adjusting the tax for inattention in later periods.

Sophisticated agents anticipate the cumulative effect of inattention on habit persistence through the δ parameter. A larger α implies a larger δ, which causes the sophisticated agent to reduce consumption during attentive periods $i = 1, ..., I$, and causes the optimal tax in those periods to decline.

We summarize the relationships of the behavioral parameters (the persistence parameter α and the inattention parameter θ) and the optimal taxes across different time periods and agent types in the following proposition:

Proposition 2. *For the fully naive household:*

- *the stronger is habit persistence, the larger the tax in all periods:*

$$\frac{\partial \tau_{tF}}{\partial \alpha} > 0, \ t = 1, ..., T.$$

- *the greater the inattention (the smaller is θ), the larger the tax in all periods in which inattention is possible. The tax during the initial period when prices are perfectly salient is unaffected by future insalience:*

$$\frac{\partial \tau_{1F}}{\partial \theta} = 0, \ \frac{\partial \tau_{tF}}{\partial \theta} < 0 \ \ for \ \ t = 2, ..., T.$$

For the partially naive household:

- *the tax in all periods is independent of the strength of habit persistence:*

$$\frac{\partial \tau_{tP}}{\partial \alpha} = 0, \ t = 1, ..., T.$$

- *the greater the inattention (the smaller is θ), the larger the tax during periods in which inattention is possible, but the tax during the perfectly salient period (period 1) is unaffected:*

$$\frac{\partial \tau_{1P}}{\partial \theta} = 0, \ \frac{\partial \tau_{tP}}{\partial \theta} < 0 \ \ for \ \ t = 2, ..., T.$$

For the sophisticated household:

- *the stronger is habit persistence (as α affects δ), the smaller the tax during periods in which attention to true prices is possible (i.e., $t = 1, ..., I$), while the tax in all future periods is unaffected:*

$$\frac{\partial \tau_{iS}}{\partial \alpha} < 0 \;\; for \;\; i = 1, ..., I, \;\; \frac{\partial \tau_{tS}}{\partial \alpha} = 0 \;\; for \;\; t = I + 1, ..., T.$$

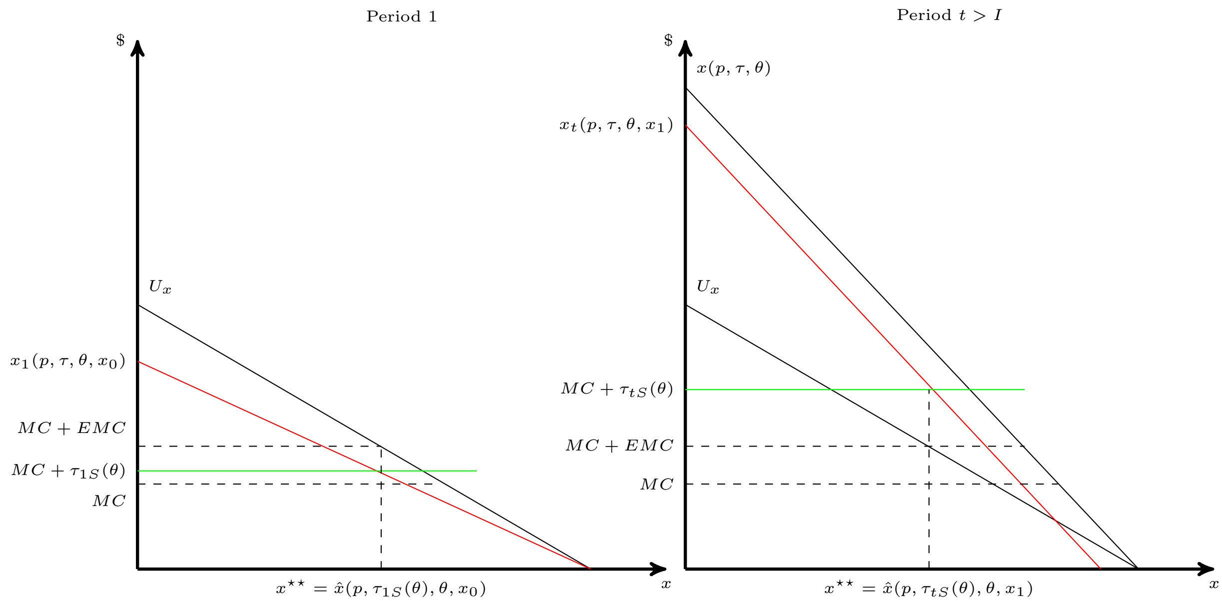
- *the greater the inattention (the smaller is θ), the larger the tax in $t > I$, and the smaller the tax during the perfectly salient period ($t = 1$). The effect on the tax during intermediate periods is ambiguous:*

$$\frac{\partial \tau_{1S}}{\partial \theta} > 0, \;\; \frac{\partial \tau_{iS}}{\partial \theta} >< 0 \;\; for \;\; i = 2, ..., I, \;\; \frac{\partial \tau_{tS}}{\partial \theta} < 0 \;\; for \;\; t = I + 1, ..., T.$$

2.3.1. Graphical Example

An example of taxes for the sophisticated agent are displayed in Figure 1. The lines labeled U_x are the privately optimal demand functions if no salience effects exist. Because prices are insalient in period $t > I$, the demand curve rotates up to $x(p, \tau, \theta)$ - more is consumed at every price, and demand is less responsive to prices. In anticipation of this the sophisticated agent consumes less in period 1. This is captured by the downward rotation to demand curve $x_1(p, \tau, \theta, x_0)$ in period 1. This reduction in consumption also reduces the habit stock inherited in period t, which shifts the period t demand curve inward to $x_t(p, \tau, \theta, x_1)$. These final demands in red are used by the agent in choosing the consumption bundle. The optimal taxes defined in Proposition 1 lead the agent to consume where there true marginal benefits given by U_x intersect the social marginal cost ($MC + EMC$).

Figure 1: Optimal Dynamic Taxes for Sophisticated Agent



2.4. Time-invariant Tax Alternatives

For a variety of pragmatic reasons, including administrative costs and political opposition, it may not be feasible to implement a time-varying tax. In this section, we explore some time-invariant tax strategies and examine the implications of those strategies for social welfare for a variety of agent types.

The most obvious candidate for a time-invariant tax is the standard Pigouvian rate of marginal external damage, ϕ. If regulators recognize that price salience is a problem, they may also apply the received wisdom from static models of tax salience (e.g., Chetty et al., 2009; Chetty, 2009). Such a tax would fail to optimally account for time-varying salience and the intertemporal nature of internalities. In this subsection we derive the optimal tax for the static model (which is suboptimal in a dynamic setting), and then we derive a second-best tax in the dynamic model that maximizes social welfare subject to the constraint that the tax is time-invariant. This second-best constant tax is a weighted average of the optimal dynamic tax in each period with weights that depend on the behavioral parameters and the internalities associated with each agent type. Subsequently, in the following subsection, we show how to quantify the welfare losses for *any* constant tax relative to the dynamic tax optimum.

2.4.1. The (Sub)optimal Static Tax

If the regulator ignores the dynamic nature of the agent's utility function, they will want to implement a consumption path that equates the marginal utility in each time period to the sum of the price and marginal external cost:

$$u_t' = p + \phi$$

The regulator will further expect agents to equate their marginal utility in each period to the perceived price:

$$u_t' = \theta p$$

The static tax chosen by a regulator would therefore equate

$$p + \phi = \theta(p + \tau_{stat})$$

The tax that solves this equation is

$$\tau_{stat} = p\frac{1-\theta}{\theta} + \frac{\phi}{\theta}. \tag{31}$$

This is equal to the optimal tax imposed on partially naive and sophisticated agents during the insalient window. Intuitively, partially naive and sophisticated agents privately optimize over their consumption persistence, so a static tax can be optimal for them during periods in which their inattention is stable. The inefficiency of the static tax arises if there are dynamic internalities (e.g., with fully naive agents) or if inattention is changing over time.

2.4.2. The Second-best Constant Tax

Let $\hat{W}(\hat{\mathbf{x}}_j, \hat{\mathbf{y}}_j) = \hat{W}_j(x_0, p; \theta)$ be the social welfare function from (20) evaluated at the privately optimal consumption paths for $j \in \{F, P, S\}$. Suppose a small tax τ is added to a previously untaxed environment so that we evaluate $\hat{W}_j(x_0, p + \tau; \theta)$. The second-best optimal constant tax solves $\frac{\partial \hat{W}}{\partial \tau} = 0$ for each type of agent. As we show below, in each case the second-best tax is an exact-weighted average of the optimal dynamic taxes where the weights depend on the household type and within-period demand curve slope. The slopes of the demand curves will be constant functions of θ, α, the price p, the habit stock at time t, and utility parameters, so for convenience define

$$\frac{\partial \hat{x}_{tj}}{\partial \tau} = a_{tj}, \quad j \in \{F, P, S\}.$$

Also for notational convenience we will denote the optimal dynamic tax τ_{tj} as $\tau_{I+1,j}$ for any period $t = I+1, ..., T$ because the optimal dynamic tax is constant during the insalient window after the $I'th$ period in each cycle, when inattention is stable.

Proposition 3. *To conserve notation, let $\delta_j = 0$ for fully and partially naive agents and $\delta_j = \delta$ for sophisticated agents. The second-best constant tax rate for $j \in \{F, P, S\}$, is given by the weighted average*

$$\tau_j^\star = w_{1j}\tau_{1j}(\theta) + \sum_{i=2}^{I} w_{ij}\tau_{ij}(\theta) + w_{I+1,j}\tau_{I+1,j}(\theta), \tag{32}$$

with weights

$$w_{1j} = \frac{a_{1j}(1+\delta_j)}{a_{1j}(1+\delta_j) + \sum_{i=2}^{I}\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta) + \theta\sum_{t=I+1}^{T}\beta^{t-1}a_{tj}},$$

$$w_{ij} = \frac{\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta)}{a_{1j}(1+\delta_j) + \sum_{i=2}^{I}\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta) + \theta\sum_{t=I+1}^{T}\beta^{t-1}a_{tj}}, \quad i = 2, ..., I,$$

$$w_{I+1,j} = \frac{\theta\sum_{t=I+1}^{T}\beta^{t-1}a_{tj}}{a_{1j}(1+\delta_j) + \sum_{i=2}^{I}\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta) + \theta\sum_{t=I+1}^{T}\beta^{t-1}a_{tj}}$$

and $\tau_{1j}(\theta)$, $\tau_{ij}(\theta)$, and $\tau_{I+1,j}(\theta)$ are defined in Proposition 1.[7]

Proof. By plugging the privately chosen consumption path for each agent[8] into the social welfare function and differentiating with respect to τ, we obtain

$$\frac{\partial \hat{W}_j}{\partial \tau} = 0 = \sum_{M=1}^{\infty}\beta^{T(M-1)}\Bigg[\sum_{i=1}^{I}\beta^{i-1}a_{ijM}\left[u'_{iM} - \alpha\beta u'_{i+1,M} - p - \phi\right] + \sum_{t=I+1}^{T}\beta^{t-1}a_{tjM}\left[u'_{tM} - \alpha\beta u'_{t+1,M} - p - \phi\right]\Bigg]$$

In the long run, behavior is repeated within each cycle so that we can drop the M subscript,

---

[7]This tax can also be more compactly expressed as

$$\tau_j^\star = \frac{\sum_{i=1}^{T}\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta)\cdot\tau_{ij}(\theta)}{\sum_{i=1}^{T}\beta^{i-1}a_{ij}(\mu_i(1+\delta_j)+(1-\mu_i)\theta)}, \quad \mu_1 = 1; \;\; \mu_i \in [0,1] \text{ for } i = 2, ..., I; \;\; \mu_i = 0 \text{ for } i > I. \tag{33}$$

[8]The privately chosen consumption path is implicit in (8) and (12) if $j = F$, implicit in (9) if $j = P$, and given in (16) and the solution to (18) if $j = S$.

and the expression becomes

$$\frac{\partial \hat{W}_j}{\partial \tau} = 0 = \frac{1}{1-\beta^T}\left[\sum_{i=1}^{I} \beta^{i-1} a_{ij} \left[u_i' - \alpha\beta u_{i+1}' - p - \phi\right] \right. \\ \left. + \sum_{t=I+1}^{T} \beta^{t-1} a_{tj} \left[u_t' - \alpha\beta u_{t+1}' - p - \phi\right]\right]$$

which is equivalent to

$$\frac{\partial \hat{W}_j}{\partial \tau} = 0 = \sum_{i=1}^{I} \beta^{i-1} a_{ij} \left[u_i' - \alpha\beta u_{i+1}' - p - \phi\right] + \sum_{t=I+1}^{T} \beta^{t-1} a_{tj} \left[u_t' - \alpha\beta u_{t+1}' - p - \phi\right]$$

If $j = F$, the first order conditions in (8) and (12) (modified to include the introduction of a small tax) can be used to simplify this to

$$0 = \sum_{i=1}^{I} \beta^{i-1} a_{iF} \left[(p+\tau)(1-\alpha\beta)(\mu_i + (1-\mu_i)\theta) - p - \phi\right] + \sum_{t=I+1}^{T} \beta^{t-1} a_{tF} \left[(p+\tau)(1-\alpha\beta)\theta - p - \phi\right]$$

Solving for τ and plugging in the formulas for $\tau_{1F}(\theta)$, $\tau_{iF}(\theta)$, and $\tau_{I+1,F}(\theta)$ gives the result.

Similarly if $j = P$, the first order conditions in (9) (again modified to include the introduction of a small tax) can be used to simplify $\frac{\partial \hat{W}_P}{\partial \tau}$ to

$$0 = \sum_{i=1}^{I} \beta^{i-1} a_{iP} \left[(p+\tau)(\mu_i + (1-\mu_i)\theta) - p - \phi\right] + \sum_{t=I+1}^{T} \beta^{t-1} a_{tP} \left[(p+\tau)\theta - p - \phi\right]$$

Solving for τ and plugging in the formulas for $\tau_{1P}(\theta)$, $\tau_{iP}(\theta)$, and $\tau_{I+1,P}(\theta)$ gives the result.

Lastly, if $j = S$, the first order conditions in (14) and (19) (again modified to include the introduction of a small tax) can be used to simplify $\frac{\partial \hat{W}_S}{\partial \tau}$ to

$$0 = \sum_{i=1}^{I} \beta^{i-1} a_{iS} \left[(p+\tau)(\mu_i(1+\delta) + (1-\mu_i)\theta) - p - \phi\right] + \sum_{t=I+1}^{T} \beta^{t-1} a_{tS} \left[\theta(p+\tau) - p - \phi\right]$$

Solving for τ and plugging in the formulas for $\tau_{1S}(\theta)$, $\tau_{iS}(\theta)$, and $\tau_{I+1,S}(\theta)$ gives the result.

■

The weights in $\tau_j^\star$ are the discounted demand curve slopes, adjusted by the salience applied to each period's decision. The second-best tax therefore differs between agent types for two reasons. First, the weights differ between agent types; the demand curves have different slopes because the partially naive agent plans for habits while the fully naive agent does not, and the sophisticated agent plans for both habits and inattention. These different dynamic behaviors create different price-responsiveness within given period. Second, the optimal dynamic taxes that are being weighted also differ by agent type.

The size of the second best constant tax for sophisticated agents is more influenced by the behavioral parameters of the habit process than the second best taxes for the naive agent types because of the sophisticated agent's demand for commitment. If consumption is more persistent (if α is large), then sophisticated agents have a more effective commitment device. This makes future demand more responsive to current consumption (i.e., the $\frac{\partial \hat{x}_t}{\partial x_i}$ are large, making δ large), and the agent makes larger corrections in early periods on her own. In this case, the constant tax must be weighted more towards correcting under-consumption in early periods. With a large δ, τ_{1S} and τ_{iS} become more heavily weighted and also become smaller, so the constant tax $\tau_S^\star$ also becomes smaller. This does not necessarily mean that stronger habit persistence is better for welfare. Intuitively, with stronger habit persistence there is more adjustment in consumption across price salient and insalient time periods. Part of the consumption- and pollution-reducing functions of the second-best constant tax are weakened in order to prevent the sophisticated agent from consuming too little when prices are salient. A large δ will also occur if the discount factor is close to one (so that the agent adjusts its habits out of concern for the future and needs less of a tax incentive). Likewise if δ is small the emphasis in the constant tax can be shifted back towards reducing pollution and overconsumption in later periods, and the tax is larger.

2.5. Efficiency Cost

We now turn to deriving the welfare loss of a time-invarient tax. The formulas we derive are valid for calculating the efficiency cost of any time-invariant tax, however we will focus primarily on the second-best constant tax $\tau_j^\star$. Although $\tau_j^\star$ is chosen in each case to minimize the deadweight loss from over- and under-consumption by definition, the remaining losses have policy relevance. The size of the deadweight loss (and of $\tau_j^\star$) depends on behavioral parameters α and θ, as well as the level of sophistication of the agent. To the extent that alternative policy tools (such as information technology investments or goal-setting programs) can affect these parameters, they can influence the size of the deadweight loss. For example, smart electricity grid technologies that provide real-time price information to households require large fixed costs, but the new information that households receive may alter their inattention and habit formation behavior. Therefore an estimable expression for this loss and how it depends on these parameters is important for these investment decisions.

We will use the concept of equivalent variation to derive an expression for the excess burden on consumers when the regulator imposes a tax that cannot correct all internalities and externalities. The loss will be calculated as the net present value of the wealth society would be willing to forgoe in order to avoid the imperfect tax instrument (relative to the optimum), net of any changes in tax revenue between the two policies. Because the reference (optimal) tax level varies, we will represent society's expenditure function and indirect welfare function as depending on a tax-inclusive price vector (even though the tax-inclusive price will be constant for a the second-best constant tax). Temporarily suppressing the F, P, and S subscripts, the general expression for the excess burden of deviating from the $I+1$-vector of optimal taxes $\underline{\tau}(\theta)$ to $\underline{\tau}^{\star}$ is

$$EB(\underline{\tau}^{\star}) = \frac{m}{1-\beta} - e\left(p+\tau_1(\theta), ..., p+\tau_{I+1}(\theta), \widehat{W}(p+\tau^{\star}, ..., p+\tau^{\star}, m, x_0; \theta)\right) \\ - \left(R(\tau^{\star}_{(1)}, ..., \tau^{\star}_{(I+1)}, m) - R(\tau_1(\theta), ..., \tau_{I+1}(\theta), m)\right) \quad (34)$$

Although $\tau^{\star}$ is a constant, in order to compare its application in different time periods with $\underline{\tau}(\theta) = (\tau_1(\theta), ..., \tau_{I+1}(\theta))$, we denote $\underline{\tau}^{\star} = \left(\tau^{\star}_{(1)}, ..., \tau^{\star}_{(I+1)}\right)$ to be the vector containing $\tau^{\star}$ applied in the analogous time periods, with $\tau^{\star}_{(I+1)}$ applied in $t = I+1, ..., T$. Note that we could use the same procedure to calculate the excess burden of any $I+1$-vector of sub-optimal taxes, include the other time-invariant tax options discussed in the previous subsection. The function $e(\cdot)$ is society's expenditure function. It is the amount of wealth at the optimal tax vector that would achieve the level of welfare obtained under the second-best constant tax. The difference between $e(\cdot)$ and the net present value of income is the amount of wealth society would be willing to forgoe to retain the optimal tax structure. The last term in parentheses measures any gain or loss in tax revenue from moving from the optimal dynamic taxes to the alternative set of taxes. For example,

$$R(\tau_1(\theta), ..., \tau_{I+1}(\theta), m) = \sum_{M=1}^{\infty} \beta^{T(M-1)} \sum_{t=1}^{T} \beta^{t-1} \tau_t(\theta) \hat{x}_{tM}(\tau_1(\theta), ..., \tau_{I+1}(\theta), m, x_0; \theta).$$

Following Auerbach (1985), we derive a more convenient expression for $EB(\underline{\tau}^{\star})$ using a second-order Taylor expansion. We calculate the expansion around the optimal dynamic tax vector $\underline{\tau}(\theta)$.

$$EB(\underline{\tau}^{\star}) \approx \frac{\partial EB}{\partial \underline{\tau}'}(\underline{\tau}(\theta)) \cdot (\underline{\tau}^{\star} - \underline{\tau}(\theta)) + \frac{1}{2}(\underline{\tau}^{\star} - \underline{\tau}(\theta))' \cdot \frac{\partial^2 EB}{\partial \underline{\tau} \partial \underline{\tau}'}(\underline{\tau}(\theta)) \cdot (\underline{\tau}^{\star} - \underline{\tau}(\theta)) \tag{35}$$

Again, we can use this expression to calculate the efficiency cost of any suboptimal tax vector by substituting that tax vector in for $\underline{\tau}^{\star}$. The envelope theorem guarantees that the first-order terms in the Taylor expansion will be zero when evaluated at the optimal tax vector. The second order terms require taking derivatives of marginal utilities in each period with respect to tax changes in the same and all other periods, evaluated at the demands associated with the optimal dynamic tax vector.

Let $P_t(x_{t-1}, x_t, x_{t+1}) = u_t'(x_t - \alpha x_{t-1}) - \alpha\beta u_{t+1}'(x_{t+1} - \alpha x_t)$ represent the agent's true intertemporal preferences for x_t, i.e., the welfare-relevant inverse demand at time t. This is the inverse demand curve at time t along the optimal path if the agent was a perfectly attentive and sophisticated agent. The decision-relevant demands will deviate from this welfare-relevant demand because of inattention and, in the case of the fully naive agent, the failure to optimize over habits.

For an optimizing agent, past and anticipated future quantities demanded shift current demand. Partially naive and sophisticated agents adjust current demand in response to anticipated future tax changes because they expect the persistence of current consumption to affect their future tax burden. During the salient window of each cycle, the sophisticated agent also anticipates being inattentive to the future tax-inclusive price; in order to avoid paying additional future taxes due to inattention, the agent further reduces consumption today. A future tax increase therefore results in an inward rotation in the early period demand curves for sophisticated agents. That inward rotation causes reduced consumption to persist through lower habit in future periods, causing future period demand curves to shift inward. Fully naive agents, on the other hand, do not anticipate future habits, so future tax changes do not affect current period demand. Current period tax increases do influence current period demand, however, which has a spillover effect on future consumption through passive habit persistence. For fully naive agents, therefore, all past tax changes - but no future tax changes - affect demand in a given period.

Consider the effect of a tax change applied in period $i \leq I$ on demand in some arbitrary period t, and suppose the tax change is applied in the $i'th$ period of each cycle. This effect will include a direct effect if $i = t$, as well as indirect effects that persist from changes in demand in the $i'th$ period of previous cycles, and indirect effects from anticipated demand changes in the $i'th$ period of future cycles if agents are partially naive or sophisticated:

$$\frac{dx_{jt,M}}{d\tau_{j,(i)}} = \begin{cases} \underbrace{\sum_{m=1}^{M-1} \frac{\partial x_{jt,M}}{\partial x_{ji,m}} \frac{\partial x_{ji,m}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects persisting from earlier} \\ \text{periods } i \text{ in previous cycles}}} + \underbrace{\frac{\partial x_{jt,M}}{\partial \tau_{j,(i)}}}_{\text{Direct effect}} + \underbrace{\sum_{n=M+1}^{\infty} \frac{\partial x_{jt,M}}{\partial x_{ji,n}} \frac{\partial x_{ji,n}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects anticipated from} \\ \text{changes in later cycles}}} & \text{for } t = i, \\ \underbrace{\sum_{m=1}^{M} \frac{\partial x_{jt,M}}{\partial x_{ji,m}} \frac{\partial x_{ji,m}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects persisting from earlier} \\ \text{periods } i \text{ in current and previous cycles}}} + \underbrace{\sum_{n=M+1}^{\infty} \frac{\partial x_{jt,M}}{\partial x_{ji,n}} \frac{\partial x_{ji,n}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects anticipated from} \\ \text{changes in later cycles}}} & \text{for } t > i, \\ \underbrace{\sum_{m=1}^{M-1} \frac{\partial x_{jt,M}}{\partial x_{ji,m}} \frac{\partial x_{ji,m}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects persisting from earlier} \\ \text{periods } i \text{ in previous cycles}}} + \underbrace{\sum_{n=M}^{\infty} \frac{\partial x_{jt,M}}{\partial x_{ji,n}} \frac{\partial x_{ji,n}}{\partial \tau_{j,(i)}}}_{\substack{\text{Indirect effects anticipated from later} \\ \text{periods } i \text{ in current and future cycles}}} & \text{for } t < i, \end{cases} \tag{36}$$

where all of the indirect effects anticipated from future changes are zero for fully naive agents.

The $I+1$'th element of the tax vector is applied in each period during the insalient window, so the direct effect on demand in each of those periods has an indirect effect on each other period in the current and all other cycles. If the period t of interest is within the insalient window in a particular cycle, the total effect of a change in the $I+1$'th element of the tax vector is

$$\begin{aligned} \frac{dx_{jt,M}}{d\tau_{j,(I+1)}} = & \underbrace{\sum_{m=1}^{M-1} \left(\frac{\partial x_{jt,M}}{\partial x_{j,I+1,m}} \frac{\partial x_{j,I+1,m}}{\partial \tau_{j,(I+1)}} + ... + \frac{\partial x_{jt,M}}{\partial x_{jT,m}} \frac{\partial x_{jT,m}}{\partial \tau_{j,(I+1)}} \right)}_{\substack{\text{Indirect effects from insalient} \\ \text{window in previous cycles}}} \\ & + \underbrace{\sum_{t'=I+1}^{t-1} \frac{\partial x_{jt,M}}{\partial x_{jt',M}} \frac{\partial x_{jt',M}}{\partial \tau_{j,(I+1)}}}_{\substack{\text{Indirect effects from past} \\ \text{insalient periods in current cycle}}} + \underbrace{\frac{\partial x_{jt}}{\partial \tau_{j,(t)}}}_{\text{Direct effect}} + \underbrace{\sum_{t''=t+1}^{T} \frac{\partial x_{jt,M}}{\partial x_{jt'',M}} \frac{\partial x_{jt'',M}}{\partial \tau_{j,(I+1)}}}_{\substack{\text{Indirect effects from later} \\ \text{insalient periods in current cycle}}} \\ & + \underbrace{\sum_{n=M+1}^{\infty} \left(\frac{\partial x_{jt,M}}{\partial x_{j,I+1,n}} \frac{\partial x_{j,I+1,n}}{\partial \tau_{j,(I+1)}} + ... + \frac{\partial x_{jt,M}}{\partial x_{jT,n}} \frac{\partial x_{jT,n}}{\partial \tau_{j,(I+1)}} \right)}_{\substack{\text{Indirect effects from insalient} \\ \text{window in future cycles}}}, \quad \text{for } t > I, \end{aligned} \tag{37}$$

If the period t of interest is before the insalient window in a particular cycle, the total effect is similar except that there is no direct effect, and all indirect effects from the current cycle occur in the future. In this case the total effect of a change in the $I+1$'th element of the tax vector is

$$\frac{dx_{jt,M}}{d\tau_{j,(I+1)}} = \underbrace{\sum_{m=1}^{M-1}\left(\frac{\partial x_{jt,M}}{\partial x_{j,I+1,m}}\frac{\partial x_{j,I+1,m}}{\partial \tau_{j,(I+1)}} + ... + \frac{\partial x_{jt,M}}{\partial x_{jT,m}}\frac{\partial x_{jT,m}}{\partial \tau_{j,(I+1)}}\right)}_{\substack{\text{Indirect effects from insalient} \\ \text{window in previous cycles}}}$$
$$+ \underbrace{\sum_{n=M}^{\infty}\left(\frac{\partial x_{jt,M}}{\partial x_{j,I+1,n}}\frac{\partial x_{j,I+1,n}}{\partial \tau_{j,(I+1)}} + ... + \frac{\partial x_{jt,M}}{\partial x_{jT,n}}\frac{\partial x_{jT,n}}{\partial \tau_{j,(I+1)}}\right)}_{\substack{\text{Indirect effects from insalient window} \\ \text{later in current and future cycles}}}, \quad \text{for } t \leq I. \quad (38)$$

where in both cases all the indirect effects from future periods and cycles are zero for fully naive agents.

Lastly, let

$$\Delta\tau_{jr} = \tau^{\star}_{j,(r)} - \tau_{jr}(\theta)$$

be the deviation between the second-best constant tax applied in period r and the optimal period r tax.

Proposition 4. *The excess burden for agent type j under the second-best tax vector $\underline{\tau}^{\star}$ instead of the optimal tax vector $\underline{\tau}(\theta)$ is given by*

$$EB_j = -\frac{1}{2}\sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}\sum_{s=1}^{I+1}\sum_{r=1}^{I+1}\Delta\tau_{js}\Delta\tau_{jr}\left(\frac{\partial P_{tM}}{\partial x_{j,t-1,M}}\frac{dx_{j,t-1,M}}{d\tau_{js}}\right.$$
$$\left. + \frac{\partial P_{tM}}{\partial x_{jtM}}\frac{dx_{jtM}}{d\tau_{js}} + \frac{\partial P_{tM}}{\partial x_{j,t+1,M}}\frac{dx_{j,t+1,M}}{d\tau_{sj}}\right)\frac{dx_{jtM}}{d\tau_{jr}}. \quad (39)$$

Proof.

Write the derivative of excess burden with respect to the r'th element of the tax vector as

$$\begin{aligned}\frac{\partial EB_j}{\partial \tau_r} &= P_1(x_0,x_1,x_2)\frac{dx_{j1}}{d\tau_r} + \beta P_2(x_1,x_2,x_3)\frac{dx_{j2}}{d\tau_r} + \dots \\ &= \sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}P_{tM}(x_{j,t-1,M},x_{jtM},x_{j,t+1,M})\frac{dx_{jtM}}{d\tau_r}.\end{aligned}$$

Taking second- and cross-partials of this expression gives

$$\frac{\partial^2 EB_j}{\partial \tau_s \partial \tau_r} = \sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}\left(\frac{\partial P_{tM}}{\partial x_{j,t-1,M}}\frac{dx_{j,t-1,M}}{d\tau_{js}} + \frac{\partial P_{tM}}{\partial x_{jtM}}\frac{dx_{jtM}}{d\tau_{js}} + \frac{\partial P_{tM}}{\partial x_{j,t+1,M}}\frac{dx_{j,t+1,M}}{d\tau_{js}}\right)\frac{dx_{jtM}}{d\tau_{jr}}.$$

Because the first order terms in the Taylor approximation to excess burden are zero at the optimal tax, we can rewrite (35) as

$$EB(\underline{\tau}^{\star}) \approx -\frac{1}{2}\left(\underline{\tau}^{\star} - \underline{\tau}(\theta)\right)' \begin{bmatrix} \frac{\partial^2 EB}{\partial \tau_1^{\star 2}} & \cdots & \frac{\partial^2 EB}{\partial \tau_1^{\star} \partial \tau_{I+1}^{\star}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 EB}{\partial \tau_1^{\star} \partial \tau_{I+1}^{\star}} & \cdots & \frac{\partial^2 EB}{\partial \tau_{I+1}^{\star 2}} \end{bmatrix} \left(\underline{\tau}^{\star} - \underline{\tau}(\theta)\right) \tag{40}$$

Expanding equation (40) and plugging in the second- and cross-partial derivatives produces equation (39).

■

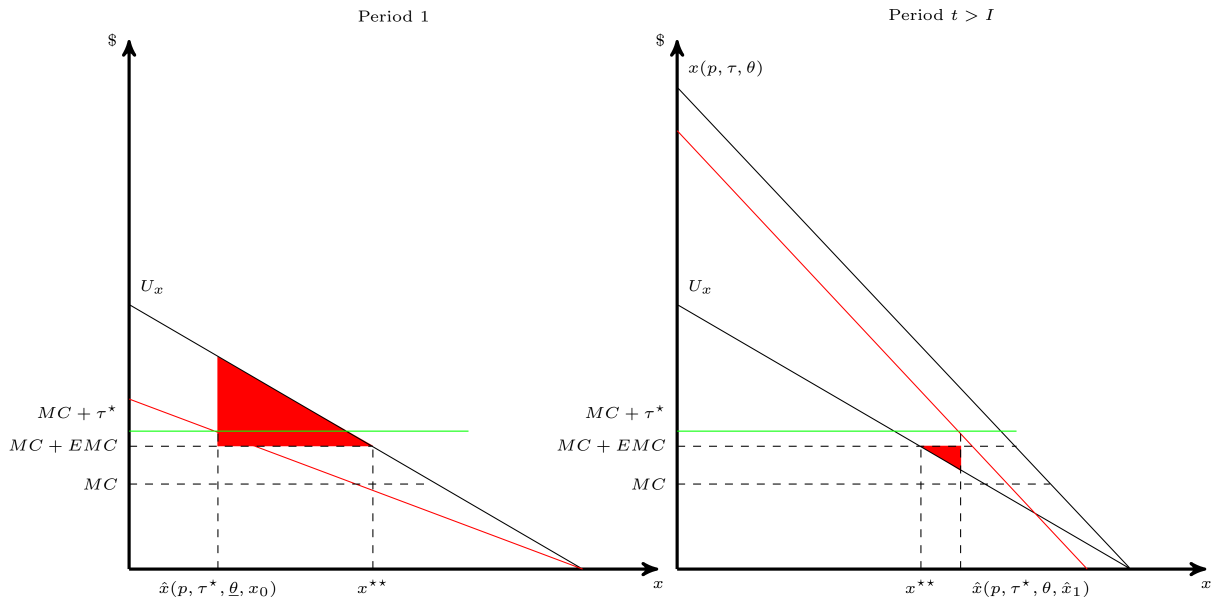
The second order terms that govern the Taylor series approximation to $EB(\underline{\tau}^{\star})$ are simply the net present value of a sequence of deadweight loss triangles. The base and height of the deadweight loss triangle in each period needs to adjust for the (potential) rotations and shifts that occur because of the response to tax changes in different periods. If there were no habits or inattention then no such shifts or rotations would occur; in this case the deadweight loss (dropping the cycle subscript M for convenience) in each period would be given by the

$$-\frac{1}{2}\Delta\tau_{jt}^2 \frac{\partial P_t}{\partial x_{jt}} \left(\frac{\partial x_{jt}}{\partial \tau_{jt}}\right)^2$$

terms in equation (39). Furthermore, during the insalient window, these deadweight loss triangles would be $-\frac{1}{2}\theta\Delta\tau_{jt}\frac{\partial x_{jt}}{\partial \tau_{jt}}\Delta\tau_{jt}$ because the ratio of the welfare-relevant demand slope to the decision-relevant demand slope, $\frac{\partial P_t}{\partial x_{jt}}\frac{\partial x_{jt}}{\partial \tau_{jt}}$, is the expected salience parameter in period t. For sophisticated households during the salient window the deadweight loss triangles would be $-\frac{1}{2}\left(\mu_t(1+\delta) + (1-\mu_t)\theta\right)\Delta\tau_{jt}\frac{\partial x_{jt}}{\partial \tau_{jt}}\Delta\tau_{jt}$. The remainder of the terms in equation (39) capture the additional shifts and rotations from the indirect effects of taxes applied in other periods that further distort the size of the deadweight loss triangles in each period.

The deadweight loss for sophisticated agents is illustrated in Figure 2 (for the first period and a later period during the insalient window).

Figure 2: Excess Burden of the Second-best Constant Tax for Sophisticated Agents



As drawn, the deadweight loss in period 1 is larger than in period $t > I$, but the period $t > I$ losses are experienced during each period in the insalient window in a given cycle. The total excess burden in (40) is increasing in the habit parameter α, holding taxes constant. When the taxes during the insalient window are lowered from their optimal dynamic values to meet the second-best constant rate, overconsumption in the insalient window becomes a problem and the sophisticated agent makes larger adjustments during the salient window, rotating the period 1 demand curve further downward. Agents with larger α will make larger period 1 adjustments. The more the demand curve rotates downward, the greater the underconsumption problem in period 1, and the greater the period 1 deadweight loss, which increases total losses.

Equation (39) is useful for quantifying the social losses more generally if some tax vector $\underline{\widetilde{\tau}}$ is implemented that differs from $\underline{\tau}_j(\theta)$. For example, we can evaluate the excess burden from any of the time-invariant taxes discussed in subsection 2.4. If the regulator uses the conventional wisdom of taxing at marginal external damage $\phi_{(s)}$ in each period s then the excess burden is simply

$$EB_{j,\phi} = -\frac{1}{2}\sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}\sum_{s=1}^{I+1}\sum_{r=1}^{I+1}(\phi_{(s)} - \tau_{js}(\theta))(\phi_{(r)} - \tau_{jr}(\theta)) \cdot \left(\frac{\partial P_{tM}}{\partial x_{j,t-1,M}}\frac{dx_{j,t-1,M}}{d\tau_s} + \frac{\partial P_{tM}}{\partial x_{jtM}}\frac{dx_{jtM}}{d\tau_s} + \frac{\partial P_{tM}}{\partial x_{j,t+1,M}}\frac{dx_{j,t+1,M}}{d\tau_s}\right)\frac{dx_{jtM}}{d\tau_r}. \quad (41)$$

The losses from implementing alternative taxes can be calculated in a similar manner. For example, if the regulator uses the tax that would be optimal in a static model of salience we could substitute $\tau_{stat,(s)}$ in for $\phi_{(s)}$ in equation (41).[9]

3. Numerical Simulation

One useful feature of our proposed corrective tax structure is that all the parameters in the expressions are estimable or recoverable from recent studies in the literature. For the residential electricity consumption example, there is a large literature on demand elasticity, as well as recent estimates of habit persistence and price salience. Marginal social damages of various pollutants from the electricity sector are available from, among others, Graff Zivin, Kotchen, and Mansur (2012). In this section we perform a numerical simulation in order to demonstrate the magnitude of the optimal dynamic and second best corrective taxes, and the excess burden of failing to implement optimal dynamic taxes.

In order to calculate numerical values for the taxes and the excess burden we need to derive specific functional forms for the derivatives of the demand functions for each agent, and for the slope of the welfare-relevant demand. The complete expressions for these total and partial derivatives, and for the excess burden, are derived in the appendix. Here we briefly outline the characteristics of the demand curves that are used in these expressions

[9] This also applies if the regulator is incorrect about agent type. Suppose households are type j but the regulator believes they are type j' and implements $\underline{\tau}^{\star}_{j'}$. We can calculate the excess burden by plugging in the demand curve slopes for type j households applied to differences between the type j' second best taxes ($\underline{\tau}^{\star}_{j'}$) and the type j optimal taxes ($\underline{\tau}_j(\theta)$) as in:

$$EB_{j,j'} = -\frac{1}{2}\sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}\sum_{s=1}^{I+1}\sum_{r=1}^{I+1}(\tau^{\star}_{j'(s)} - \tau_{js}(\theta))(\tau^{\star}_{j'(r)} - \tau_{jr}(\theta)) \cdot \left(\frac{\partial P_{tM}}{\partial x_{j,t-1,M}}\frac{dx_{j,t-1,M}}{d\tau_s} + \frac{\partial P_{tM}}{\partial x_{jtM}}\frac{dx_{jtM}}{d\tau_s} + \frac{\partial P_{tM}}{\partial x_{j,t+1,M}}\frac{dx_{j,t+1,M}}{d\tau_s}\right)\frac{dx_{jtM}}{d\tau_r}. \quad (42)$$

before discussing the calibration and results. We consider a version of the model in which utility is quadratic. Let utility in period s be given by

$$\begin{aligned} U_s &= u(x_s - \alpha x_{s-1}) + y_s \\ &= a(x_s - \alpha x_{s-1}) - \frac{1}{2} b(x_s - \alpha x_{s-1})^2 + y_s. \end{aligned}$$

Beginning from some arbitrary period t, we derive the welfare-relevant inverse demands, $P_t(x_{t-1}, x_t, x_{t+1})$, by solving for the privately optimal plan that maximizes

$$\sum_{s=t}^{\infty} \beta^{s-t} U_s$$

subject to the budget constraint

$$A = \sum_{s=t}^{\infty} \beta^{s-t}(p x_s + y_s) \tag{43}$$

with x_{t-1} given and prices fully salient across time. The first order conditions give

$$\begin{aligned} P(x_{t-1}, x_t, x_{t+1}) &= a - b(x_t - \alpha x_{t-1}) - \beta\alpha(a - b(x_{t+1} - \alpha x_t)) \\ &= a(1 - \beta\alpha) - b(1 + \beta\alpha^2)x_t + b\alpha x_{t-1} + b\beta\alpha x_{t+1}, \end{aligned}$$

We also need to derive expressions for $\frac{dx_{jt}}{d\tau_s}$ for each agent type j and period s. These total derivatives require expressions for

- $\frac{\partial \hat{x}_{jt}}{\partial \tau}$, the slope of the type j demand curve in each period t, and
- $\frac{\partial \hat{x}_{jt}}{\partial x_{jr}}$, the shift in the type j demand curve in period t because of a change in consumption in period r that persists from the past or is anticipated in the future.

In the appendix, we derive these expressions using the demand curves for each agent type. The fully naive agent is not forward-looking about consumption persistence and ignores the intertemporal effect on habits. Rearranging the first order condition for the fully naive agent gives the following linear representation of demand in period t:

$$x_t = \frac{a}{b} - \frac{\mu_t + (1 - \mu_t)\theta}{b} p + \alpha x_{t-1}. \tag{44}$$

Equation (44) gives the familiar result that consumption is less responsive to prices when they are not fully salient. If prices are constant, then insalient prices, i.e. periods when θ is less than one, imply an upward shift in the intercept of the consumption function during the insalient window, which was the finding in Gilbert and Graff Zivin (2014)[10]. Ignoring the salience factor, this specification is equivalent to the case in Becker, Grossman, and Murphy (1994) in which habit formation is not forward-looking. If there is no consumption persistence ($\alpha = 0$) then current consumption also does not depend directly on past consumption.

Rearranging the first order condition for the problem for partially naive and sophisticated agents, who are forward-looking about consumption persistence, provides the following linear demand:

$$x_{tj} = \frac{a(1-\beta\alpha)}{b(1+\beta\alpha^2)} - \frac{(\mu_t(1+\delta_j) + (1-\mu_t)\theta)}{b(1+\beta\alpha^2)} p + \frac{\alpha}{1+\beta\alpha^2} x_{t-1} + \beta \frac{\alpha}{1+\beta\alpha^2} x_{t+1}, \tag{45}$$

where $\delta_j = 0$ for partially naive agents and $\delta_j = \delta$ for sophisticated agents.

In a model without consumption persistence or inattention, the response to a price change in the current period is $-1/b$. This is dampened by $1/(1+\beta\alpha^2)$ as agents anticipate the persistence of today's consumption change into the future. In the case of partially naive agents, the response to current prices is also dampened by the expected salience $\mu_t + (1-\mu_t)\theta$. In the case of sophisticated agents, the response to current prices is augmented by $1 + \delta$ as the household tries to commit to reduced consumption.

3.1. Numerical Calibration

We first provide parameters for a central case of the model for U.S. electricity consumption, and then perform sensitivity analysis for key parameters. The parameters for the central case are given in Table 1.

The habit parameter can be calculated from empirical studies on habit formation in energy consumption. Filippini et al. (2016) estimate a coefficient on lagged electricity consumption of 0.422. Using the derivations above, we solve $0.422 = \frac{\alpha}{1+\beta\alpha^2}$ for α, using the root that falls between zero and one[11]. The daily discount factor β was chosen to produce an annual discount factor of 0.9.

[10] We assume a constant price in order to isolate the effect of price salience rather than price uncertainty. This is also consistent with our empirical context because electricity prices are usually fixed, regulated rates.

[11] Using Scott's (2012) coefficient of 0.787 on lagged gasoline consumption and solving for the real part of the complex roots produces $\alpha = 0.64$, while using Heien & Durham's (1991) coefficient of 0.418 on lagged electricity produces $\alpha = 0.54$. Macroeconomic studies using aggregate data (e.g., Fuhrer (2000)) typically estimate the habit parameter closer to 0.8.

The salience parameter is calculated using Sexton's (2015) estimate that households that pay their electricity bill using automatic withdrawal consume 4% more energy than houses that receive a bill and take an action to make a payment. We take the automatic withdrawal households as inattentive to price, and the active payment households as attentive to price. Using a demand elasticity of -0.1, the implied relative slope of the inattentive and attentive demand curves is 0.71. The astute reader will note that this estimate is larger than those found in the tax salience literature (e.g., Chetty et al. (2009) and Taubinsky & Rees-Jones (2017)), but those studies assume that only the tax portion of the tax-inclusive price is insalient. We will evaluate sensitivity of our results to this parameter below.

The marginal damage parameter is calculated by combining an estimate of the national average marginal CO_2/kwh emissions with an estimate of the marginal social cost of CO_2. Graff Zivin et al. (2014) estimate a national average marginal emissions rate of 1.21 lbs of CO_2/kwh. Inflated to 2017 dollars, the U.S. EPA's most recent central estimate of the social cost of carbon is \$49.66 per ton of CO_2 which we converted to pounds to arrive at an estimate of \$0.03/kwh, or about one fourth of the 2017 national average electricity price of \$0.129/kwh (EIA, 2017).[12]

We use the price elasticity of electricity demand, along with the national average daily household consumption and electricity price, in order to calculate an average linear demand slope. We assume that electricity prices are not salient for the majority of each billing month (Gilbert & Graff Zivin (2014)), and so we take our calculated slopes to represent the slope during insalient periods, despite the fact that they "average in" brief periods of salience following payment of a bill. Estimates of the elasticity of electricity demand vary in the literature. We do not wish to provide a full review of the electricity demand elasticity literature here, but as a few recent examples, Alberini & Filippini (2011) estimate short run elasticities between -0.8 and -0.15, Filippini et al. (2016) between -0.12 and -0.27, and Deryugina et al. (2017) of about -0.16. We take the low end of this range and use -0.1 because we have very short run (daily) behavior in mind for our model.

Lastly, we assume a cycle length of $T = 30$ to reflect monthly billing, and a salient window of $I = 10$ which is roughly consistent with the findings in Gilbert and Graff Zivin (2014). The magnitudes of our results are stable over variation in the length of the salient window within a cycle.

[12]EPA (2013) reported the social cost of carbon as \$42 in the year 2020 (valued in 2007 dollars) at a 3% social discount rate. We converted this to 2017 dollars using the CPI.

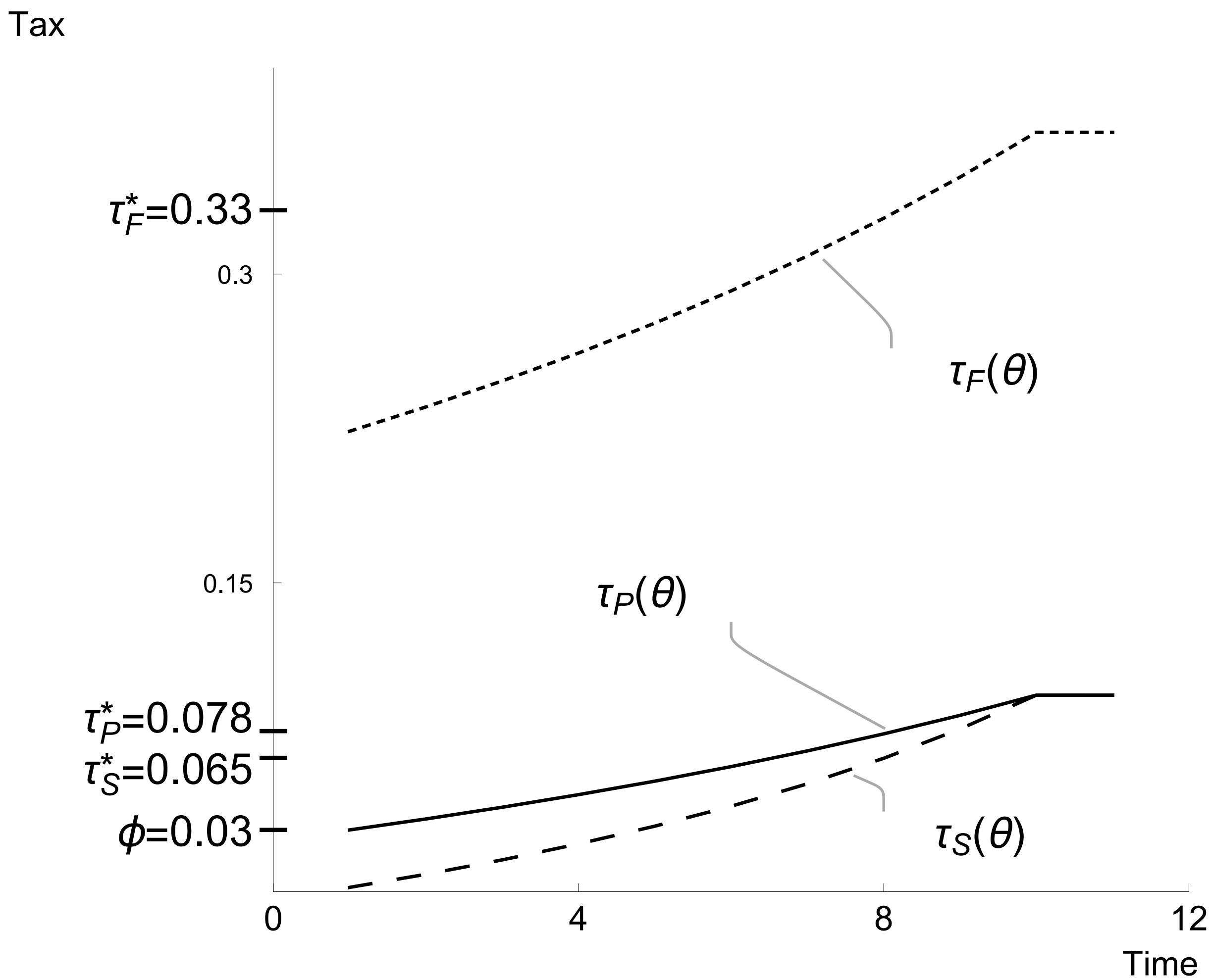
Table 1: Calibration Parameters

Parameter		Value
Discount factor	β	0.9997
Habit	α	0.55
Salience	θ	0.71
Marginal damage	ϕ	\$0.03/kwh
Electricity price	p	\$0.129/kwh
Average consumption	$\overline{x}$	29.5 kwh/day
Demand elasticity	$\epsilon = \frac{dx_t}{dp}\frac{p}{\overline{x}}$	-0.1
Cycle length	T	30
Salient window	I	10

3.2. Results

The time path of optimal dynamic taxes ($\tau_j(\theta)$) for each agent type is shown in Figure 3, with second-best constant taxes ($\tau_j^\star$) marked on the vertical axis. The taxes are shown in dollars per unit of electricity, or kilowatthour. For reference, the U.S. average electricity price is $0.129 per kilowatthour. As stated in Proposition 1, the taxes are rising throughout the salient window as the probability of full price attention declines. At any time period, the tax is significantly larger for the fully naive agent than either of the other agent types or the marginal externality of $0.03 per kilowatthour. The tax for the partially naive agent begins at the marginal externality in the first, fully salient period, and is always at least as large as the tax for the sophisticated agent. The tax for the sophisticated agent begins below the marginal externality and rises to exceed it by the end of the salient window.

Figure 3: Optimal dynamic and second-best taxes



Notes: The dotted line is the optimal dynamic tax for the Fully Naive agent, the solid black line is the optimal dynamic tax for the Partially Naive agent, and the dashed line is the optimal dynamic tax for the Sophisticated agent. Bold hash marks on the vertical axis denote second-best constant taxes in each case, as well as the marginal social damage, $\phi = 0.03$. Tax units are in dollars; for example, ϕ is three cents per kilowatthour of electricity which is about one fourth of the national average electricity price of 12.9 cents per kilowatthour.

The main results for excess burden are shown in Table 2, including the numerical values of the taxes shown in Figure 3. The excess burden of a time invariant tax is reported for the second-best constant tax, the static salience tax ignoring dynamic inefficiencies ($\tau_{stat} = 0.095$), and the standard Pigouvian rate of $\phi = 0.03$. Excess burden is reported in present value dollars. The excess burden of the second-best constant tax is largest for the fully naive agent at \$114.5 per household and smallest for the partially naive household at \$18.93 per household. With approximately 122 million households in the U.S., this puts the welfare loss associated with moving from the optimal dynamic tax to the optimal, but second-best, constant tax at between \$2.4 billion and \$14.4 billion for the residential electricity sector. If agents are sophisticated, this figure is \$49.91 per household or \$6.3 billion for the U.S. residential electricity sector.

Notice, however, that the excess burden increases significantly as the time-invariant tax moves away from it's second-best optimal level. When taxing at the static optimal rate of \$0.095, which ignores dynamic inefficiencies, the excess burden is \$2,364, \$30.28 or \$90.61 per household if agents are fully naive, partially naive, or sophisticated, respectively. Whe taxing at the Pigouvian rate of \$0.03, which ignores internalities, the excess burden is \$3,790, \$84.84, or \$83.90 per household if agents are fully naive, partially naive, or sophisticated, respectively. Table 2 also reports excess burden figures for implementing no taxes at all, which are not surprisingly larger than any of the tax scenarios discussed.

Table 2: Excess burden and tax results for the base case

Taxes	Fully Naive	Partially Naive	Sophisticated
Excess burden of time-invariant taxes in present value $			
$EB(\tau_j^\star)$	114.5	18.93	49.91
$EB(\tau_{stat})$	2,364	30.28	90.61
$EB(\phi)$	3,790	84.84	83.90
Excess burden when there is no tax			
$EB(0)$	4,561	198.4	181.4
Second-best constant taxes in $ per kwh			
$\tau_j^\star$	0.331	0.078	0.065
Optimal dynamic taxes in $ per kwh			
$\tau_1(\theta)$	0.224	0.030	0.002
$\tau_2(\theta)$	0.235	0.035	0.008
$\tau_3(\theta)$	0.248	0.041	0.015
$\tau_4(\theta)$	0.262	0.047	0.023
$\tau_5(\theta)$	0.276	0.054	0.032
$\tau_6(\theta)$	0.292	0.061	0.041
$\tau_7(\theta)$	0.309	0.068	0.052
$\tau_8(\theta)$	0.327	0.077	0.065
$\tau_9(\theta)$	0.347	0.086	0.079
$\tau_{10}(\theta)$	0.369	0.095	0.095
$\tau_{I+1}(\theta)$	0.369	0.095	0.095

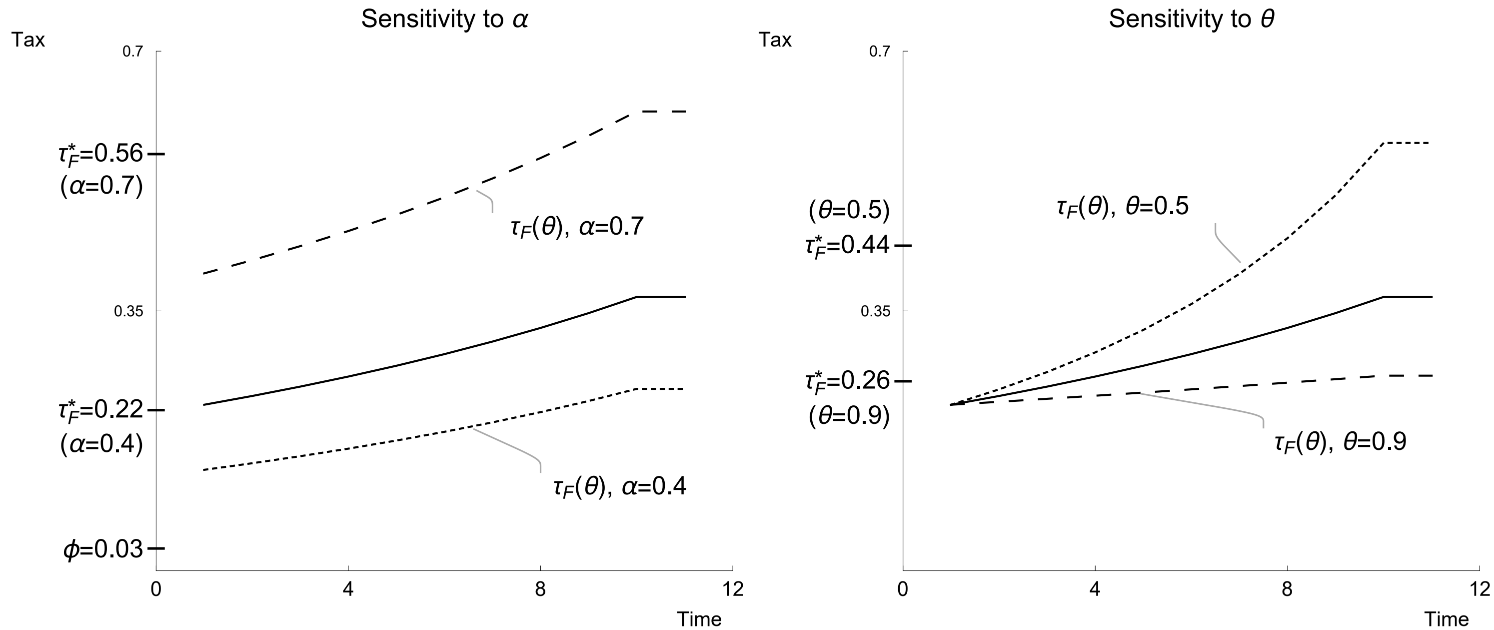
Notes: This table reports the optimal dynamic taxes, second-best constant taxes, and excess burden of several time-invariant taxes using parameters from Table 1. The taxes are reported in dollars per unit of electricity (kilowatthour) while the excess burden values are reported in present value dollars. The marginal externality is $\phi = \$0.03$ per unit of electricity (kilowatthour) and the optimal tax for a static model is $\tau_{stat} = \$0.095$ per kilowatthour. For comparison, the U.S. average residential electricity price is $0.129 per kilowatthour.

Figure 4 illustrates the findings of Proposition 2, on the sensitivity of optimal dynamic taxes to persistence (α) and salience (θ). For the fully naive agent, stronger habit persistence requires a larger tax in all periods because the household fails to take consumption persistence into account. The partially naive agent, on the other hand, privately optimizes its habit stock in each period, so the tax only needs to correct for the salience internality and the externality; as a result the habit parameter has no effect on the size of the tax. The sophisticated agent also privately optimizes its habit stock, except that during attentive periods the agent is imperfectly committing to a new consumption path through habit persistence and the regulator must optimally address that commitment demand. If commitment (i.e., habit) is stronger, the regulator can reduce the tax during potentially attentive periods.

For all three agent types, the tax can be lower during inattentive periods if inattention is less severe (i.e., if θ is closer to one). During attentive periods the relationship between θ and the optimal tax differs across agent types. The fully naive and partially naive agents do not anticipate inattention in the future, so θ only affects the tax during periods in which there is a chance the agent will be inattentive. The sophisticated agent underconsumes in early periods because of (imperfect) commitment to reduce future overconsumption. Because of early underconsumption, the tax starts below marginal damage in the first, perfectly salient period and rises to exceed marginal damage as the probability of being inattentive rises. The tax in early periods can therefore be larger (closer to marginal damage) if future inattention is less severe, as this alleviates underconsumption during the early periods.

Figure 4: Sensitivity of optimal taxes to persistence α and salience θ, Continued on next page...

(a) Fully Naive



(b) Partially Naive

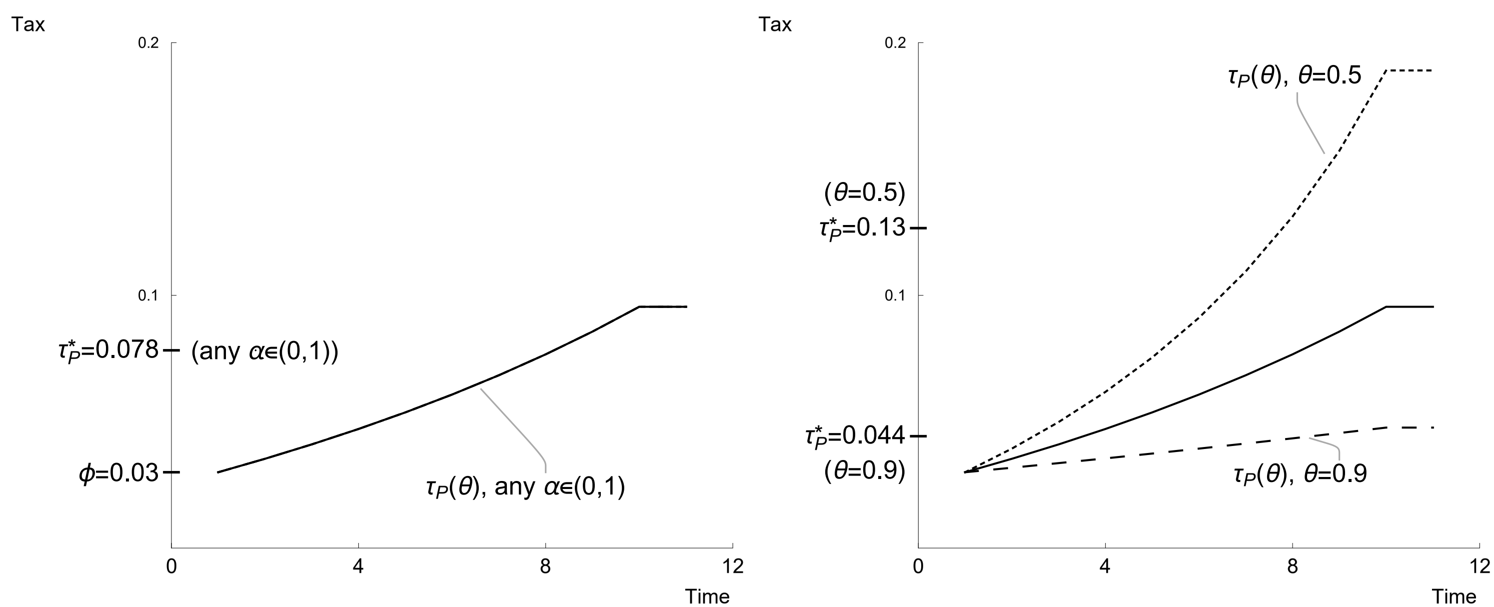
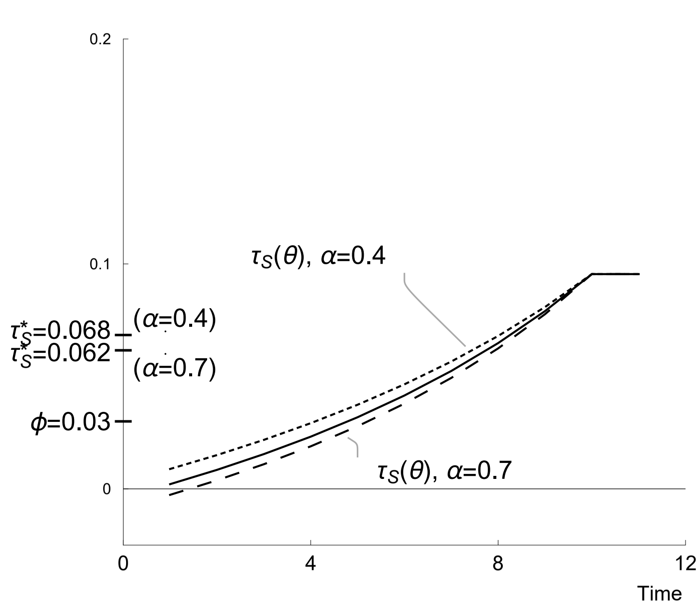
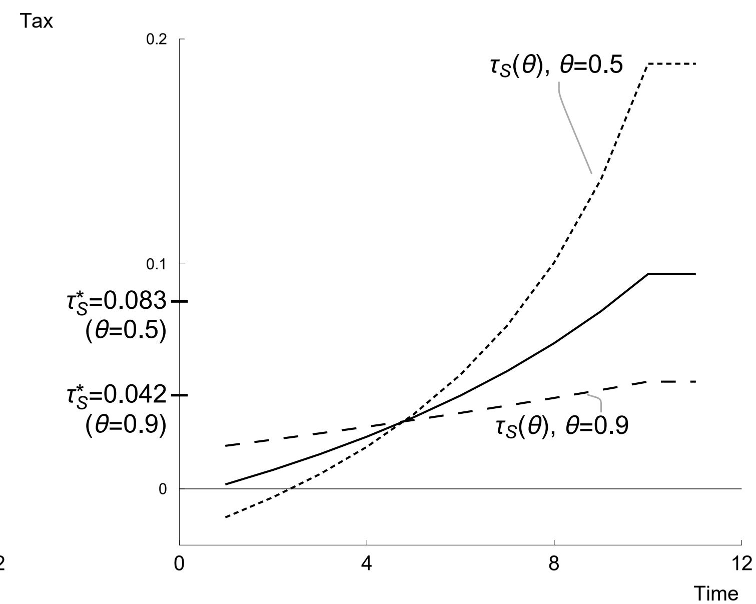


Figure 4: ... continued from previous page.

(c) Sophisticated



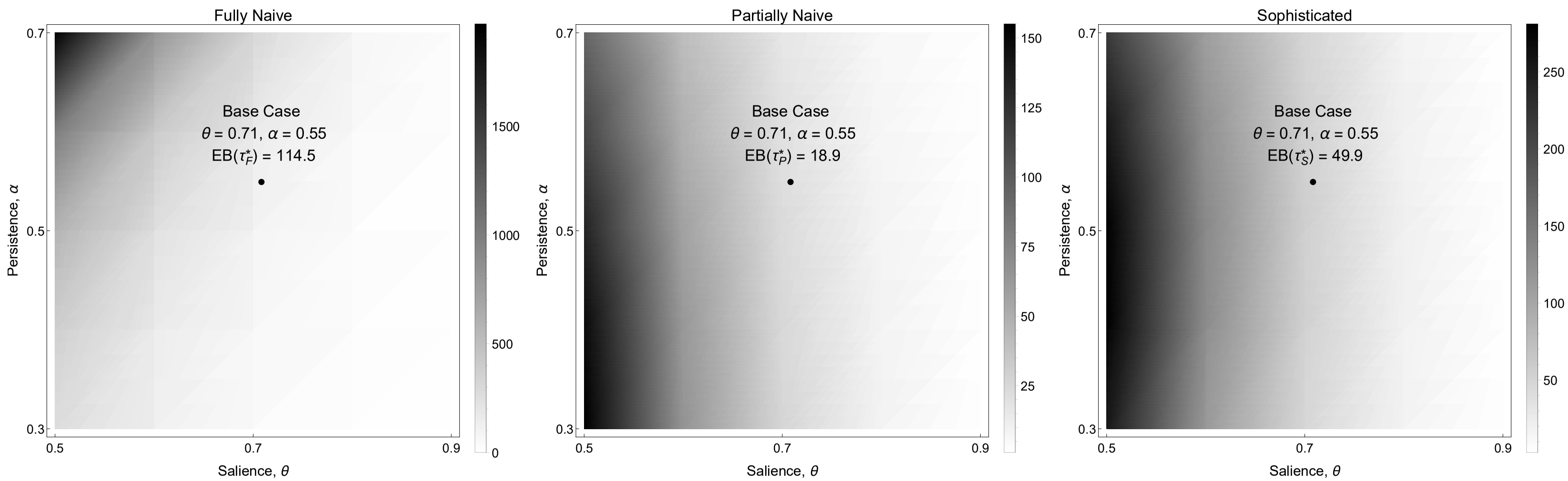


Notes: The solid black line on each graph is the optimal dynamic tax in the base case, with $\alpha = 0.55$ and $\theta = 0.71$. Bold hash marks on the vertical axes denote second-best constant taxes in each case, as well as the marginal social damage, $\phi = 0.03$. Tax units are in dollars; for example, ϕ is three cents per kilowatthour of electricity which is about one fourth of the national average electricity price of 12.9 cents per kilowatthour. Note the different scale of the vertical axis for the Fully Naive case. For Fully Naive agents, optimal taxes are larger in all periods with greater consumption persistence, α, and with lower salience, θ (other than the first period). For Partially Naive agents, optimal taxes are invariant to α but are larger with lower θ. For Sophisticated agents, optimal taxes are lower during the insalient window when α is larger, whereas a lower θ's causes the time path of optimal taxes to rotate down in the earliest periods and up in later periods.

Figure 5 illustrates the sensitivity of excess burden to changes in the habit persistence and salience parameters. Darker shading corresponds to larger excess burden. The scale of the subfigures differs for each agent type because the excess burden increases so rapidly as the salience parameter declines, but at different rates for each agent type; it is difficult to represent sensitivity to θ on the same scale. For fully naive agents, the excess burden of the second-best tax increases rapidly both as salience declines and persistence increases. This occurs because myopia about habit persistence creates a large internality for the fully naive agent and as the persistence parameter increases, this internality grows. For partially naive and sophisticated agents, the major source of variation in excess burden is the salience parameter, because this is the main source of the internality. However, there is a slight decline in excess burden for the partially naive agent as persistence increases. In Figures 6 and 7 of Appendix B, we report similar results for the excess burden of the static salience tax and the Pigouvian tax.

Lastly, Table 3 reports the sensitivity of excess burden to the elasticity of demand. Specifically, the table shows that the excess burden for each of the time-invariant taxes and each of the agent types grows linearly as demand becomes more elastic. The welfare losses for a good that has unit elastic demand are ten times larger than the welfare losses in our residential electricity example assuming a demand elasticity of -0.1. Considering the variety of types of goods to which our model applies (e.g., gasoline, unhealthy food, etc.), the economy-wide losses from imperfect taxation and time-varying salience are likely to be non-trivial.

Figure 5: Excess burden of second-best constant tax ($\tau = \tau^{\star}$)
Sensitivity to persistence α and salience θ



Notes: Excess burden is quantified in per household present value dollars, with darker shading indicating larger excess burden. Note that the scale is different for each agent type. This is necessary because the variation in excess burden across the parameter space and agent types is highly nonlinear. For Fully Naive agents, both lower salience and higher persistence have an economically meaningful effect on the excess burden of a constant tax. For Partially Naive and Sophisticated agents, who dynamically optimize over persistence, salience is the main factor in the size of the excess burden of a constant tax. However, excess burden does vary slightly with persistence; excess burden is larger for when persistence is lower for Partially Naive agents, whereas excess burden is concave in persistence for Sophisticated agents.

Table 3: The sensitivity of excess burden to demand elasticity

	Fully Naive			Partially Naive			Sophisticated		
ϵ	$EB_F(\tau_F^\star)$	$EB_F(\phi)$	$EB_F(\tau_{stat})$	$EB_P(\tau_P^\star)$	$EB_P(\phi)$	$EB_P(\tau_{stat})$	$EB_S(\tau_S^\star)$	$EB_S(\phi)$	$EB_S(\tau_{stat})$
−0.1	114	3790	2360	18.9	84.8	30.3	49.9	83.9	90.6
−0.2	229	7580	4730	37.9	170	60.6	99.8	168	181
−0.3	343	11400	7090	56.8	255	90.8	150	252	272
−0.4	458	15200	9460	75.7	339	121	200	336	362
−0.5	572	19000	11800	94.6	424	151	250	419	453
−0.6	687	22700	14200	114	509	182	299	503	544
−0.7	801	26500	16500	132	594	212	349	587	634
−0.8	916	30300	18900	151	679	242	399	671	725
−0.9	1030	34100	21300	170	764	272	449	755	815
−1.0	1140	37900	23600	189	848	303	499	839	906
−1.1	1260	41700	26000	208	933	333	549	923	997
−1.2	1370	45500	28400	227	1020	363	599	1010	1090
−1.3	1490	49300	30700	246	1100	394	649	1090	1180
−1.4	1600	53100	33100	265	1190	424	699	1170	1270
−1.5	1720	56900	35500	284	1270	454	749	1260	1360

Notes: This table reports the excess burden of time invariant taxes as a function of the elasticity of demand, ϵ. Three time-invariant taxes are evaluated for each agent type: the second best constant tax ($\tau_j^\star$), the Pigouvian tax which ignores internalities (ϕ), and the static salience tax (τ_{stat}). In each case, excess burden scales linearly with the elasticity of demand.

4. Conclusion

Economies across the globe are becoming increasingly cashless and many payments systems have become automated, driving a temporal wedge between consumption and payment and generally making the costs of consumption intermittently salient. Since this inconsistent price salience alters demand elasticities, it is a particular concern for goods that generate externalities and the price-based policies deployed to address them. This paper develops a simple model of consumer behavior when prices of a good that creates some social harm are intermittently salient. Dynamics are driven by the persistence of consumption decisions across time, an empirical feature of intertemporal consumption for which there is a large volume of evidence across contexts such as habit formation, durable goods adoption, and status quo bias, to name a few. In this setting the forward-looking capacity of the agents shapes the nature of dynamic externalities and internalities, and is therefore crucial for optimal taxation and welfare analysis.

We derive an optimal dynamic tax schedule for three types of agents: a *fully naive* agent who does not anticipate future inattention or plan for consumption persistence; a *partially naive* agent who also does not anticipate future inattention but is forward looking about persistence; and a *sophisticated* agent who anticipates future inattention and plans for consumption persistence. In the fully naive case, optimal taxes are highly sensitive to the degree of salience and persistence, with lower salience and higher persistence requiring considerably larger taxes to generate the optimal consumption path. In the partially naive case, the degree of persistence is irrelevant to the optimal tax schedule because there is no internality from persistence and also no demand for commitment. Lower salience still results in larger optimal taxes, however. In the sophisticated case, the agent demands commitment to avoid future suboptimal decisions when prices are salient, and consumption persistence provides a vehicle to at least partially achieve that commitment. Greater persistence therefore lowers the taxes required to put the agent on the optimal consumption path, while lower salience results in greater variation in the tax schedule: lower taxes when agents are paying attention to prices, and higher optimal taxes when they are not.

Since time-varying taxes can be bureaucratically expensive and politically difficult to implement, we also explore the consumption and welfare implications under a variety of time-invariant tax strategies. In particular, we derive a second-best constant tax that maximizes welfare among time-invariant taxes. We also consider the welfare consequences of ignoring internalities and taxing at the Pigouvian rate of marginal external cost, as well as the welfare consequences of ignoring dynamics and imposing a tax that would be optimal in a static model with inattention and externalities.

The key insights from our model are then placed into context through a simulation exercise based on data from the U.S. residential electricity market. This market is attractive for a number of reasons, not least of which because there is well-documented persistence in consumption over time within households, price inattention, intermittent billing, and socially costly environmental externalities. Our base case results suggest that optimal taxes are significantly larger than those corresponding to the simple Pigouvian case. The magnitude of this difference is driven by the size of the internalities associated with inattention and persistence and vary by agent type. The optimal tax is roughly ten times larger than the Pigouvian one for fully naive agents, but only twice as large for sophisticated agents. We also find that the welfare losses from time-invariant taxes are more than an order of magnitude larger if agents are fully naive than if agents are partially naive or sophisticated. Even the optimal time-invariant tax generates welfare losses that range from 2.4 to 14.5 billion dollars when aggregated across the US electricity market.

The sensitivity of these results to alternative parameter assumptions is also explored. Since the excess burden from second-best tax strategies scales with the elasticity of demand, the welfare losses in this context will very much depend on the nature of demand for the good in question. Goods with a unit elasticity of demand, for example, would incur welfare losses ten times that of our electricity example. The degree of price salience also plays a significant role in determining the welfare losses from second best time-invariant tax policies. A modest thirty percent change in the salience parameter generates an excess burden that is five to ten times larger than in our base case, underscoring the magnitude of heterogenous impacts across different consumption contexts.

Our study has several limitations. Our model of inattention is intentionally stylized. We do not allow price inattention to be determined by past decisions, the magnitude of relative prices, or other environmental cues that may determine attention in a particular context. While our conceptualization of consumption persistence is general, it abstracts from important aspects of persistence that could arise due to complementary durable good purchases. Efforts to endogenize inattention and further formalize the persistence relationship constitute the next logical steps in advancing the insights from our model. On the empirical side, our simulation is hampered by the lack of data and empirical tests that could be used to distinguish between fully naive, partially naive, and sophisticated agents. Understanding which agent types are most prevalent in a particular setting, or better yet how to target each of them through careful mechanism design, is critical for the design of good policy and an area ripe for future research.

Appendix A

In this appendix we derive an explicit expression for excess burden in the case of quadratic utility. In order to apply a specific functional form, it will be convenient to rewrite the expression for excess burden from Proposition 4 as

$$EB_j = -\frac{1}{2}\sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}\sum_{s=1}^{I+1}\sum_{r=1}^{I+1}\Delta\tau_{js}\Delta\tau_{jr} \cdot u''_{tM}\left(\frac{dx_{jtM}}{d\tau_{js}} - \alpha\frac{dx_{j,t-1,M}}{d\tau_{js}}\right)\left(\frac{dx_{jtM}}{d\tau_{jr}} - \alpha\frac{dx_{j,t-1,M}}{d\tau_{jr}}\right). \quad (46)$$

To see that this is equivalent to the expression in Proposition 4, rewrite the derivative of the excess burden with respect to the r'th element of the tax vector as

$$\begin{aligned}\frac{\partial EB_j}{\partial\tau_r} &= P_1(x_0, x_1, x_2)\frac{dx_{j1}}{d\tau_r} + \beta P_2(x_1, x_2, x_3)\frac{dx_{j2}}{d\tau_r} + \ldots \\ &= (u'_1 - \beta\alpha u'_2)\frac{dx_{j1}}{d\tau_r} + \beta(u'_2 - \beta\alpha u'_3)\frac{dx_{j2}}{d\tau_r} + \ldots \\ &= u'_1\left(\frac{dx_{j1}}{d\tau_r} - \alpha\cdot 0\right) + \beta u'_2\left(\frac{dx_{j2}}{d\tau_r} - \alpha\frac{dx_{j1}}{d\tau_r}\right) + \ldots \\ &= \sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}u'_{tM}\left(\frac{dx_{jtM}}{d\tau_{jr}} - \alpha\frac{dx_{j,t-1,M}}{d\tau_{jr}}\right).\end{aligned}$$

Taking second- and cross-partials of this expression gives

$$\frac{\partial^2 EB_j}{\partial\tau_s\partial\tau_r} = \sum_{M=1}^{\infty}\sum_{t=1}^{T}\beta^{T(M-1)+t-1}u''_{tM}\left(\frac{dx_{jtM}}{d\tau_{js}} - \alpha\frac{dx_{j,t-1,M}}{d\tau_{js}}\right)\left(\frac{dx_{jtM}}{d\tau_{jr}} - \alpha\frac{dx_{j,t-1,M}}{d\tau_{jr}}\right).$$

In the quadratic case, conveniently $u''_t = -b$. We now need to derive expressions for $\left(\frac{dx_{jt}}{d\tau_s} - \alpha\frac{dx_{j,t-1}}{d\tau_s}\right)$ for each j and s. As mentioned in the text, these expressions will utilize

- $\frac{\partial\hat{x}_{jt}}{\partial\tau}$, the slope of the type j demand curve in each period t, and
- $\frac{\partial\hat{x}_{jt}}{\partial x_{jr}}$, the shift in the type j demand curve in period t because of a change in consumption in period r that persists from the past or is anticipated in the future.

4.1. Fully Naive Agents

As discussed in the text, fully naive agents with quadratic utility have the following linear demand curve:

$$x_t = \frac{a}{b} - \frac{\mu_t + (1-\mu_t)\theta}{b} p + \alpha x_{t-1}. \tag{47}$$

The period $t+1$ demand depends on how much was consumed in period t through the habit process such that $\hat{x}_{t+1} = \hat{x}_{t+1}(x_t, p; \theta)$, with $\hat{x}_{t+1}$ increasing in x_t. As before, $\frac{\partial \hat{x}_{t+1}}{\partial x_t} > 0$ is a constant. In this case,

$$\frac{\partial \hat{x}_{t+1}}{\partial x_t} = \alpha, \quad \frac{\partial \hat{x}_{t+2}}{\partial x_t} = \alpha^2, \quad \text{... etc.}$$

Recall that the $I + 1'th$ element of the dynamic tax vector is imposed in every period during the insalient window, and that for fully naive agents the effect of a future tax increase on current demand is zero. From equation (44), we obtain for tax changes during the salient window

$$\frac{dx_{tM}}{d\tau_r} = \begin{cases} -\frac{\mu_r + (1-\mu_r)\theta}{b} \sum_{m=1}^{M} \alpha^{T(M-m)+t-r} & \text{if } r \leq I, \quad r \leq t, \\ -\frac{\mu_r + (1-\mu_r)\theta}{b} \sum_{m=1}^{M-1} \alpha^{T(M-m)+t-r} & \text{if } r \leq I, \quad r > t. \end{cases} \tag{48}$$

The slope of the demand curve in period r when the tax is levied governs the period r change in x_r, which persists forward to period t according to α^{t-r}, α^{T+t-r}, and so forth. Tax changes that occur later than period t do not affect period t demand.

For the tax imposed during the insalient window from $I+1$ to T, we have

$$\frac{dx_t}{d\tau_{(I+1)}} = \begin{cases} -\frac{\theta}{b}\left(\sum_{m=1}^{M-1} \sum_{t'=1}^{T-I} \alpha^{T(M-m)+t-(I+t')} + \sum_{t'=1}^{t-I} \alpha^{t-(I+t')} \right) & \text{if } I < t, \\ -\frac{\theta}{b}\left(\sum_{m=1}^{M-1} \sum_{t'=1}^{T-I} \alpha^{T(M-m)+t-(I+t')} \right) & \text{if } I \geq t. \end{cases} \tag{49}$$

This tax directly affects demand in period $I+1$, $I+2$, etc., during the current and previous cycles and these effects accumulate up to period $t > I$ (or accumulate from previous cycles only if $t \leq I$).

From these expressions it is simple to show that for $r \leq I$,

$$\frac{dx_t}{d\tau_r} - \alpha \frac{dx_{t-1}}{d\tau_r} = \begin{cases} -\frac{\mu_r + (1-\mu_r)\theta}{b} & \text{if } r = t, \\ 0 & \text{if } r \neq t. \end{cases} \tag{50}$$

and

$$\frac{dx_t}{d\tau_{(I+1)}} - \alpha \frac{dx_{t-1}}{d\tau_{(I+1)}} = \begin{cases} -\frac{\theta}{b} & \text{if } I < t, \\ 0 & \text{if } I \geq t. \end{cases} \tag{51}$$

This analysis demonstrates that $\left(\frac{dx_{jt}}{d\tau_s} - \alpha \frac{dx_{j,t-1}}{d\tau_s} \right) = 0$ unless $s = t$ for $s \leq I$, or $s = I+1$

for $t > I$. This also implies that

$$\frac{\partial^2 EB_F}{\partial \tau_s \partial \tau_r} = 0 \text{ unless } s = r.$$

For example,

$$\begin{aligned}
\frac{\partial^2 EB_F}{\partial \tau_r^2} &= \frac{\beta^{r-1}}{1-\beta^T}(-b)\left(-\frac{\mu_r+(1-\mu_r)\theta}{b}\right)^2 \text{ if } r \leq I, \\
\frac{\partial^2 EB_F}{\partial \tau_{(I+1)}^2} &= \sum_{M=1}^{\infty}\sum_{t=I+1}^{T}\beta^{T(M-1)+t-1}(-b)(-\tfrac{\theta}{b})^2 = \frac{\sum_{t=I+1}^{T}\beta^{t-1}}{1-\beta^T}(-b)(-\tfrac{\theta}{b})^2.
\end{aligned} \tag{52}$$

We can now write the excess burden for fully naive agents as

$$EB_F(\underline{\tau}^{\star}) \approx -\frac{1}{2}\left(\Delta\underline{\tau}_F\right)' \cdot$$
$$\begin{bmatrix}
-\frac{1}{b}\frac{1}{1-\beta^T} & 0 & \cdots & & 0 \\
0 & -\frac{1}{b}\frac{\beta}{1-\beta^T}\left(\mu_2+(1-\mu_2)\theta\right)^2 & & & \vdots \\
\vdots & & \ddots & & \\
 & & & -\frac{1}{b}\frac{\beta^{I-1}}{1-\beta^T}\left(\mu_I+(1-\mu_I)\theta\right)^2 & 0 \\
0 & \cdots & & 0 & -\frac{1}{b}\frac{\sum_{t=I+1}^{T}\beta^{t-1}}{1-\beta^T}(\theta)^2
\end{bmatrix} \cdot (\Delta\underline{\tau}_F) \tag{53}$$

4.2. Partially Naive and Sophisticated Agents

As discussed in the text, partially naive and sophisticated agents have the following linear demand:

$$x_{tj} = \frac{a(1-\beta\alpha)}{b(1+\beta\alpha^2)} - \frac{(\mu_t(1+\delta_j)+(1-\mu_t)\theta)}{b(1+\beta\alpha^2)}p + \frac{\alpha}{1+\beta\alpha^2}x_{t-1} + \beta\frac{\alpha}{1+\beta\alpha^2}x_{t+1}, \tag{54}$$

where $\delta_j = 0$ for partially naive agents and $\delta_j = \delta$ for sophisticated agents.

For notational convenience, define the response to past consumption as

$$\gamma = \frac{\alpha}{1+\beta\alpha^2}$$

and the absolute value of the slope of current period demand as

$$d'_{jt} = \left| \frac{\partial x_{jt}}{\partial \tau} \right| = \frac{(\mu_t(1+\delta_j)+(1-\mu_t)\theta)}{b(1+\beta\alpha^2)}. \tag{55}$$

Unlike the fully naive case, current consumption is increasing in expected future consumption as well as past consumption. If the agent is attentive to prices in period t - but not sophisticated about inattention and salience - then they take into account how period t consumption affects period $t+1$ decisions, but not that period $t+1$ decisions will be suboptimal. In the sophisticated case the household also considers expected impacts on future demand through δ. Recall that

$$\delta = (1-\theta)\sum_{t=r+1}^{\infty}\beta^{t-r-1}\frac{\partial \hat{x}_t}{\partial x_r}$$

As before, $\frac{\partial \tilde{x}_{t+1}}{\partial x_t} > 0$ is a constant, but in this case,

$$\frac{\partial \tilde{x}_{t+1}}{\partial x_t} = \gamma, \quad \frac{\partial \tilde{x}_{t+2}}{\partial x_t} = \gamma^2, \quad \text{... etc.,}$$

whereas

$$\frac{\partial \tilde{x}_{t}}{\partial x_{t+1}} = \beta\gamma, \quad \frac{\partial \tilde{x}_{t}}{\partial x_{t+2}} = \beta^2\gamma^2, \quad \text{... etc..}$$

We can therefore rewrite

$$\delta = (1-\theta)\sum_{t=r+1}^{\infty}\beta^{t-r-1}\gamma^{t-r} = \frac{(1-\theta)\gamma}{1-\beta\gamma}$$

From equation (45) we obtain

$$\frac{dx_{jt}}{d\tau_r} = \begin{cases} -d'_{jr}\Big(\gamma^{t-r}\sum_{m=1}^{M}\gamma^{T(m-1)} + (\beta\gamma)^{r-t}\sum_{m=M+1}^{\infty}(\beta\gamma)^{T(m-M)}\Big) & \text{if } r \le I, \;\; r \le t, \\ -d'_{jr}\Big(\gamma^{t-r}\sum_{m=1}^{M-1}\gamma^{T\cdot m} + (\beta\gamma)^{r-t}\sum_{m=M}^{\infty}(\beta\gamma)^{T(m-M)}\Big) & \text{if } r \le I, \;\; r \ge t. \end{cases} \tag{56}$$

which can be simplified to

$$\frac{dx_{jt}}{d\tau_r} = \begin{cases} -d'_{jr}\Big(\frac{\gamma^{t-r}(1-\gamma^{T\cdot M})}{1-\gamma^T} + \frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T}\Big) & \text{if } r \le I, \;\; r \le t, \\ -d'_{jr}\Big(\frac{\gamma^{t-r+T}(1-\gamma^{T(M-1)})}{1-\gamma^T} + \frac{(\beta\gamma)^{r-t}}{1-(\beta\gamma)^T}\Big) & \text{if } r \le I, \;\; r \ge t. \end{cases} \tag{57}$$

Also from equation (45) we can obtain

$$\frac{dx_t}{d\tau_{(I+1)}} = \begin{cases} -\frac{\theta}{b(1+\beta\alpha^2)}\Big(\frac{\gamma^t(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma} + \frac{(\beta\gamma)^{T+I+1-t}}{1-(\beta\gamma)^T}\frac{1-(\beta\gamma)^{T-I}}{1-\beta\gamma} \\ \qquad\qquad + \frac{\gamma(1-\gamma^{t-I-1})}{1-\gamma} + \frac{1-(\beta\gamma)^{T-t+1}}{1-\beta\gamma}\Big) & \text{if } I < t, \\ -\frac{\theta}{b(1+\beta\alpha^2)}\Big(\frac{\gamma^t(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma} + \frac{(\beta\gamma)^{I+1-t}}{1-(\beta\gamma)^T}\frac{1-(\beta\gamma)^{T-I}}{1-\beta\gamma}\Big) & \text{if } I \geq t. \end{cases} \tag{58}$$

From these expressions algebra shows that for $r \leq I$,

$$\frac{dx_{jt}}{d\tau_r} - \alpha\frac{dx_{j,t-1}}{d\tau_r} = \begin{cases} -d'_{jr}\Big(\frac{\gamma^{1-r+T}(1-\gamma^{T(M-1)})}{1-\gamma^T} + \frac{(\beta\gamma)^{r-1}}{1-(\beta\gamma)^T}\Big) & \text{if } r \geq 1, \ \ t = 1, \\ -d'_{jr}\Big((\gamma-\alpha)\frac{\gamma^{t-1-r+T}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t}}{1-(\beta\gamma)^T}\Big) & \text{if } r \geq t > 1, \\ -d'_{jr}\Big((\gamma-\alpha)\frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T}\Big) & \text{if } r < t, \ \ r \geq 1. \end{cases} \tag{59}$$

and also that

$$\frac{dx_t}{d\tau_{(I+1)}} - \alpha\frac{dx_{t-1}}{d\tau_{(I+1)}} = \begin{cases} -\frac{\theta}{b(1+\beta\alpha^2)}\Big((1-\alpha)\Big(\frac{1}{1-\beta\gamma} + \frac{\gamma}{1-\gamma}\Big) + \frac{\alpha-\gamma}{1-\gamma}\gamma^{t-I-1} \\ \qquad + (\gamma-\alpha)\frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma} + \frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^{T+1-t}((\beta\gamma)^I-1)}{1-(\beta\gamma)^T}\Big) & \text{if } I+1 < t, \\ -\frac{\theta}{b(1+\beta\alpha^2)}\Big((\gamma-\alpha)\frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma} + \frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^{I+1-t}(1-(\beta\gamma)^{T-I})}{1-(\beta\gamma)^T}\Big) & \text{if } I+1 \geq t. \end{cases} \tag{60}$$

We can now express the excess burden for partially naive and sophisticated households. The expression can be written

$$EB_j(\underline{\tau}^\star) \approx -\frac{1}{2}\left(\Delta\underline{\tau}_j\right)' \cdot \begin{bmatrix} & \vdots & \\ \dots & \frac{\partial^2 EB_j}{\partial\tau_s\partial\tau_r} & \dots \\ & \vdots & \end{bmatrix} \cdot \left(\Delta\underline{\tau}_j\right) \tag{61}$$

where

$$\frac{\partial^2 EB_j}{\partial \tau_r^2} = -b\left(-d'_{jr}\right)^2 \sum_{M=1}^{\infty} \beta^{T(M-1)} \Bigg[\Big(\frac{\gamma^{1-r+T}(1-\gamma^{T(M-1)})}{1-\gamma^T} + \frac{(\beta\gamma)^{r-1}}{1-(\beta\gamma)^T} \Big)^2$$
$$+ \sum_{t=2}^{r} \beta^{t-1} \left((\gamma-\alpha) \frac{\gamma^{t-1-r+T}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma) \frac{(\beta\gamma)^{r-t}}{1-(\beta\gamma)^T} \right)^2$$
$$+ \sum_{t=r+1}^{T} \beta^{t-1} \left((\gamma-\alpha) \frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma) \frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T} \right)^2 \Bigg] \text{ if } r \leq I, \quad (62)$$

$$\frac{\partial^2 EB_P}{\partial \tau_{(I+1)}^2} = -b\left(-\frac{\theta}{b(1+\beta\alpha^2)}\right)^2 \times$$
$$\Bigg\{ \sum_{M=1}^{\infty} \beta^{T(M-1)} \Bigg[\sum_{t=1}^{I+1} \beta^{t-1} \Bigg((\gamma-\alpha) \frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T} \frac{1-\gamma^{T-I}}{1-\gamma} + \frac{1-\alpha\beta\gamma}{1-\beta\gamma} \frac{(\beta\gamma)^{I+1-t}(1-(\beta\gamma)^{T-I})}{1-(\beta\gamma)^T} \Bigg)^2$$
$$+ \sum_{t=I+2}^{T} \beta^{t-1} \Bigg((1-\alpha) \Big(\frac{1}{1-\beta\gamma} + \frac{\gamma}{1-\gamma} \Big) + \frac{\alpha-\gamma}{1-\gamma} \gamma^{t-I-1} + (\gamma-\alpha) \frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T} \frac{1-\gamma^{T-I}}{1-\gamma}$$
$$+ \frac{1-\alpha\beta\gamma}{1-\beta\gamma} \frac{(\beta\gamma)^{T+1-t}((\beta\gamma)^I - 1)}{1-(\beta\gamma)^T} \Bigg)^2 \Bigg] \Bigg\}, \quad (63)$$

for $r < s < I+1$,

$$\frac{\partial^2 EB_j}{\partial \tau_r \partial \tau_s} = -b(-d'_{js})(-d'_{jr})\times$$

$$\left\{ \sum_{M=1}^{\infty} \beta^{T(M-1)} \left[\left(\frac{\gamma^{1-r+T}(1-\gamma^{T(M-1)})}{1-\gamma^T} + \frac{(\beta\gamma)^{r-1}}{1-(\beta\gamma)^T} \right) \left(\frac{\gamma^{1-s+T}(1-\gamma^{T(M-1)})}{1-\gamma^T} + \frac{(\beta\gamma)^{s-1}}{1-(\beta\gamma)^T} \right) \right.\right.$$

$$+ \sum_{t=2}^{r} \beta^{t-1} \left((\gamma-\alpha)\frac{\gamma^{t-1-r+T}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t}}{1-(\beta\gamma)^T} \right) \times$$

$$\left((\gamma-\alpha)\frac{\gamma^{t-1-s+T}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{s-t}}{1-(\beta\gamma)^T} \right)$$

$$+ \sum_{t=r+1}^{s} \beta^{t-1} \left((\gamma-\alpha)\frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T} \right) \times$$

$$\left((\gamma-\alpha)\frac{\gamma^{t-1-s+T}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{s-t}}{1-(\beta\gamma)^T} \right)$$

$$+ \sum_{t=s+1}^{T} \beta^{t-1} \left((\gamma-\alpha)\frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T} \right) \times$$

$$\left.\left. \left((\gamma-\alpha)\frac{\gamma^{t-1-s}(1-\gamma^{T\cdot M})}{1-\gamma^T} + (1-\alpha\beta\gamma)\frac{(\beta\gamma)^{s-t+T}}{1-(\beta\gamma)^T} \right) \right] \right\}, \quad (64)$$

and finally

$$\frac{\partial^2 EB_j}{\partial \tau_r \partial \tau_{(I+1)}} = -b(-d'_{jr})\left(-\frac{\theta}{b(1+\beta\alpha^2)}\right)\times$$

$$\Bigg\{\sum_{M=1}^{\infty}\beta^{T(M-1)}\Bigg[\left(\frac{\gamma^{1-r+T}(1-\gamma^{T(M-1)})}{1-\gamma^T}+\frac{(\beta\gamma)^{r-1}}{1-(\beta\gamma)^T}\right)\times$$

$$\left((\gamma-\alpha)\frac{(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma}+\frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^I(1-(\beta\gamma)^{T-I})}{1-(\beta\gamma)^T}\right)$$

$$+\sum_{t=2}^{r}\beta^{t-1}\left((\gamma-\alpha)\frac{\gamma^{t-1-r+T}(1-\gamma^{T\cdot M})}{1-\gamma^T}+(1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t}}{1-(\beta\gamma)^T}\right)\times$$

$$\left((\gamma-\alpha)\frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma}+\frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^{I+1-t}(1-(\beta\gamma)^{T-I})}{1-(\beta\gamma)^T}\right)$$

$$+\sum_{t=r+1}^{I+1}\beta^{t-1}\left((\gamma-\alpha)\frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T}+(1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T}\right)\times$$

$$\left((\gamma-\alpha)\frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma}+\frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^{I+1-t}(1-(\beta\gamma)^{T-I})}{1-(\beta\gamma)^T}\right)$$

$$+\sum_{t=I+2}^{T}\beta^{t-1}\left((\gamma-\alpha)\frac{\gamma^{t-1-r}(1-\gamma^{T\cdot M})}{1-\gamma^T}+(1-\alpha\beta\gamma)\frac{(\beta\gamma)^{r-t+T}}{1-(\beta\gamma)^T}\right)\times$$

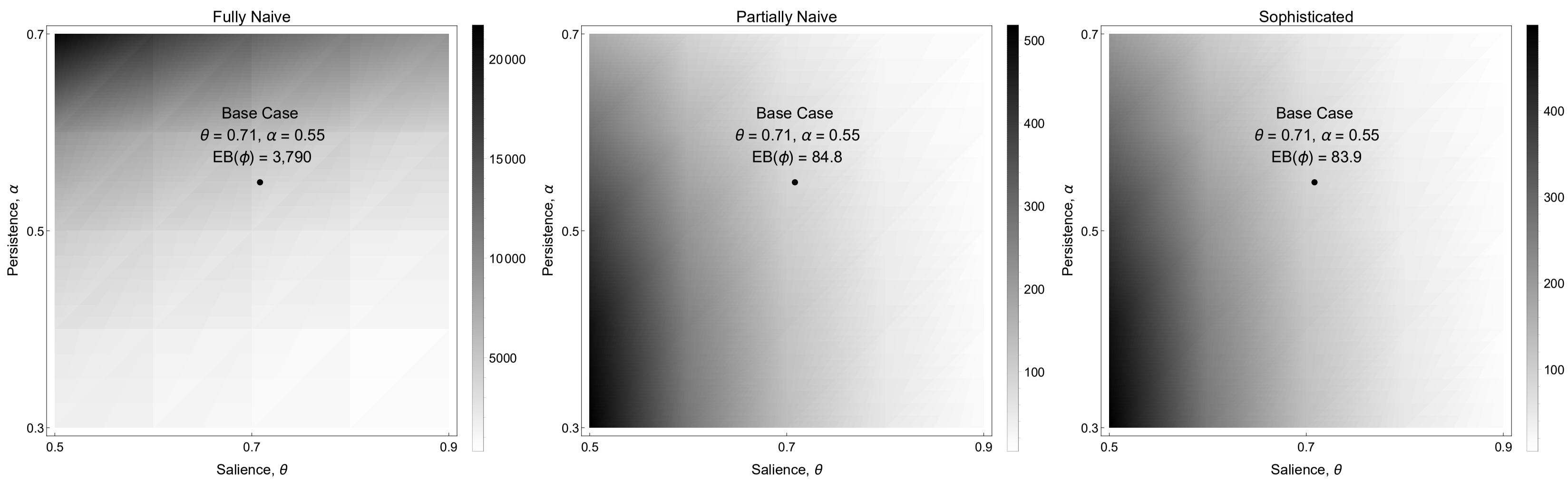
$$\Bigg((1-\alpha)\Big(\frac{1}{1-\beta\gamma}+\frac{\gamma}{1-\gamma}\Big)+\frac{\alpha-\gamma}{1-\gamma}\gamma^{t-I-1}+(\gamma-\alpha)\frac{\gamma^{t-1}(1-\gamma^{T(M-1)})}{1-\gamma^T}\frac{1-\gamma^{T-I}}{1-\gamma}$$

$$+\frac{1-\alpha\beta\gamma}{1-\beta\gamma}\frac{(\beta\gamma)^{T+1-t}((\beta\gamma)^I-1)}{1-(\beta\gamma)^T}\Bigg)\Bigg]\Bigg\}. \quad (65)$$

We now have all the expressions we need to parameterize and calculate the optimal taxes, second-best taxes, and excess burden for each household type.

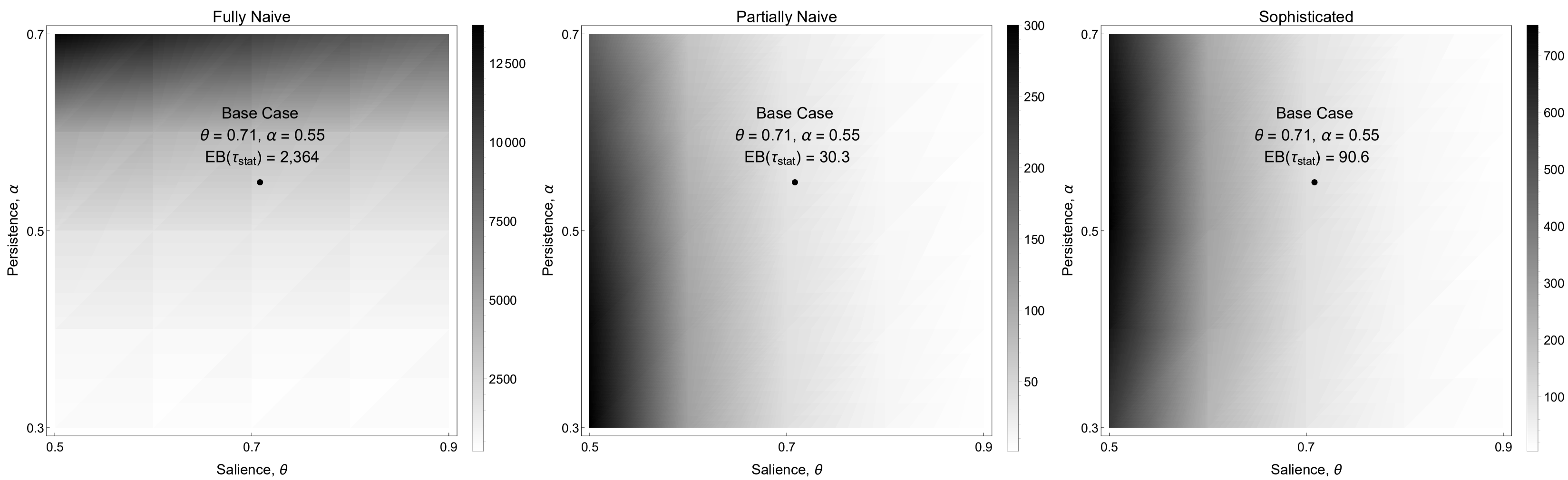
Appendix B

Figure 6: Excess burden of standard Pigouvian tax ($\tau = \phi$)
Sensitivity to persistence α and salience θ



Notes: Excess burden is quantified in per household present value dollars, with darker shading indicating larger excess burden. Note that the scale is different for each agent type. This is necessary because the variation in excess burden across the parameter space and agent types is highly nonlinear. For Fully Naive agents, both lower salience and higher persistence have an economically meaningful effect on the excess burden of a constant tax. For Partially Naive and Sophisticated agents, who dynamically optimize over persistence, salience is the main factor in the size of the excess burden of a constant tax. However, excess burden does vary slightly with persistence for these agent types; the excess burden of taxing at the standard Pigouvian rate is larger for when persistence is lower.

Figure 7: Excess burden of static salience tax ($\tau = \tau_{stat}$)
Sensitivity to persistence α and salience θ



Notes: Excess burden is quantified in per household present value dollars, with darker shading indicating larger excess burden. Note that the scale is different for each agent type. This is necessary because the variation in excess burden across the parameter space and agent types is highly nonlinear. For Fully Naive agents, both lower salience and higher persistence have an economically meaningful effect on the excess burden of a constant tax. For Partially Naive and Sophisticated agents, who dynamically optimize over persistence, salience is the main factor in the size of the excess burden of a constant tax. However, excess burden does vary slightly with persistence; excess burden is larger for when persistence is lower for Partially Naive agents, whereas excess burden is concave in persistence for Sophisticated agents.

References

Alberini, A., & Filippini, M. (2011). Response of residential electricity demand to price: The effect of measurement error. *Energy Economics, 33*(5), 889-895.

Allcott, H., & Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review, 105*(8). 2501-38.

Allcott, H., Mullainathan S., & Taubinsky, D. (2014). Energy policy with externalities and internalities. *Journal of Public Economics, 112*, 72-88.

Auerbach, A. J., (1985). The theory of excess burden and optimal taxation. In *Handbook of Public Economics* (Vol. 1, pp. 61-127). Elsevier.

Becker, G.S., & Murphy, K. M. (1988) A theory of rational addiction. *Journal of Political Economy, 96*(4), 675-700.

Becker, G. S., Grossman, M., & Murphy, K. M. (1994). An Empirical Analysis of Cigarette Addiction. The American Economic Review, 84(3), 396-418.

Benhabib, J., & Bisin, A. (2005). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumptionsaving decisions. *Games and Economic Behavior, 52*(2), 460-492.

Chetty, R. (2009). The simple economics of salience and taxation. (No. w15246). National Bureau of Economic Research.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review, 99*(4), 1145-77.

Costa, F., & Gerard, F. (2018). Hysteresis and the welfare effect of corrective policies: Theory and evidence from an energy-saving program. (No. w24608). National Bureau of Economic Research.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature, 47*(2), 315-372.

Deryugina, T., MacKay, A., & Reif, J. (2017). The long-run dynamics of electricity demand: Evidence from municipal aggregation. (No. w23483). National Bureau of Economic Research.

Farhi, E., & Gabaix, X. (2018). Optimal taxation with behavioral agents. (No. w21524). National Bureau of Economic Research.

Filippini, M., Hirl, B., & Masiero, G. (2016). Rational habits in residential electricity demand. (No. 16/228). CER-ETH Center of Economic Research at ETH Zurich, Working paper.

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

Flavin, M., & Nakagawa, S. (2008). A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence. *American Economic Review, 98*(1), 474-95.

Fuhrer, J. (2000). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review, 90*(3), 367-390.

Gilbert, B. & Graff Zivin, J. (2014). Dynamic salience with intermittent billing: Evidence from smart electricity meters. *Journal of Economic Behavior & Organization, 107*(A), 176-190.

Graff Zivin, J., Kotchen, M., & Mansur, E. (2014). Spatial and temporal heterogeneity of marginal emissions: Implications for electric cars and other electricity-shifting policies. *Journal of Economic Behavior & Organization, 107*(A), 248-268.

Grubb, M. D., & Osborne, M. (2015). Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review, 105*(1), 234-71.

Gruber, J., & Koszegi, B. (2004). Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes. *Journal of Public Economics, 88*(9-10), 1959-1987.

Heien, D., & Durham, C. (1991). A test of the habit formation hypothesis using household data. *The Review of Economics and Statistics*, 189-199.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62*(12), 3393-3411.

Landry, P. (2018). Bad habits and the endogenous timing of urges. *Review of Economic Studies, 0*, 1-22.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review, 89*(1), 103-124.

O'Donoghue, T., & Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics, 90*(10-11), 1825-1849.

Pigou, A. C. (1920). The Economics of Welfare. London: Macmillan.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*(1), 7-59.

Scott, K. R. (2012). Rational habits in gasoline demand. *Energy Economics, 34*(5), 1713-1723.

Sexton, S. (2015). Automatic bill payment and salience effects: Evidence from electricity consumption. *Review of Economics and Statistics, 97*(2), 229-241.

Taubinsky, D., & Rees-Jones, A., Forthcoming. Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies*.

U.S. Energy Information Administration (2017). Electric Power Annual 2017. U.S. Department of Energy.

U.S. Environmental Protection Agency (2013). Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866. (May 2013, Revised August 2016).