NBER WORKING PAPER SERIES

ATTENTION VARIATION AND WELFARE: THEORY AND EVIDENCE FROM A TAX SALIENCE EXPERIMENT

Dmitry Taubinsky
Alex Rees-Jones

Working Paper 22545
http://www.nber.org/papers/w22545

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016

We thank Hunt Allcott, Eduardo Azevedo, Doug Bernheim, Raj Chetty, Stefano DellaVigna, Emmanuel Farhi, Xavier Gabaix, Jacob Goldin, Tatiana Homonoff, Shachar Kariv, Supreet Kaur, Judd Kessler, David Laibson, Erzo Luttmer, Matthew Rabin, Emmanuel Saez, Andrei Shleifer, Jeremy Tobacman, Glen Weyl, Michael Woodford, Danny Yagan, and audiences at the AEA Annual Meetings, Berkeley, Carnegie Mellon SDS, the CESifo Behavioral Economics Meeting, Columbia, Cornell, Dartmouth, Haas (marketing), Harvard, the National Tax Association Annual Meetings, Stanford, Wharton, Yale for helpful comments and suggestions. We thank James Perkins at ClearVoice research for help in managing the data collection, Jessica Holevar for able research assistance, and Sargent Shriver and Vincent Conley for technical support. For financial support, we are grateful to the Lab for Economic Applications and Policy (LEAP), the Pension Research Council/Boettner Center for Pension and Retirement Research, the Russell Sage Foundation (small grants program), and the Wharton Dean's Research Fund. The opinions expressed in this paper are solely the authors', and do not necessarily reflect the views of any individual or institution listed above. The opinions expressed in this paper are solely the authors' and do not necessarily reflect the views of any individual or institution, nor those of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Dmitry Taubinsky and Alex Rees-Jones. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment
Dmitry Taubinsky and Alex Rees-Jones
NBER Working Paper No. 22545
August 2016, Revised November 2016
JEL No. C9,D0,H0

ABSTRACT

This paper shows that accounting for variation in mistakes can be crucial for welfare analysis. Focusing on consumer underreaction to not-fully-salient sales taxes, we show theoretically that the efficiency costs of taxation are amplified by 1) individual differences in underreaction and 2) the degree to which attention is increasing with the size of the tax rate. To empirically assess the importance of these issues, we implement an online shopping experiment in which 2,998 consumers—matching the U.S. adult population on key demographics—purchase common household products, facing tax rates that vary in size and salience. We find that: 1) there are significant individual differences in underreaction to taxes. Accounting for this heterogeneity increases the efficiency cost of taxation estimates by at least 200%, as compared to estimates generated from a representative agent model. 2) Tripling existing sales tax rates roughly doubles consumers' attention to taxes. Our results provide new insights into the mechanisms and determinants of boundedly rational processing of not-fully-salient incentives, and our general approach provides a framework for robust behavioral welfare analysis.

Dmitry Taubinsky
Department of Economics
Dartmouth College
Rockefeller Hall
Hanover, NH 03755
and NBER
dtaubinsky@gmail.com

Alex Rees-Jones
Operations, Information, and Decisions Department
The Wharton School
University of Pennsylvania
553 Jon M. Huntsman Hall
3730 Walnut St.
Philadelphia, PA 19104
alre@wharton.upenn.edu

1 Introduction

When incentive schemes are complex, or when certain attributes of a decision are not fully salient, consumers may make mistakes. A growing body of work documents inattention to, or incorrect beliefs about, financial incentives such as sales taxes (Chetty, Looney and Kroft, 2009), shipping and handling charges (Hossain and Morgan, 2006), energy prices (Allcott, 2015), and out-of-pocket insurance costs (Abaluck and Gruber, 2011). Such studies typically estimate the "average mistake," usually because inferring mistakes at the individual level is difficult or impossible with available data. Correspondingly, policy analysis building on these results often studies a representative agent committing the "average mistake," and thus assumes that mistakes are homogeneous.

In this paper, we demonstrate that accounting for variation in mistakes can substantially impact policy analysis. We highlight two crucial ways in which variation in mistakes matters. First, the variation in mistakes across consumers matters: the greater the individual differences, the lower the allocational efficiency of the market, because these differences drive a wedge between who buys the product and who benefits from it the most. Second, the variation in mistakes across incentive levels matters: this variation creates a debiasing channel that can accentuate the demand response to policy changes. In the theoretical component of this paper, we formalize the role of these two channels in shaping the efficiency cost of taxation when consumers misreact to sales taxes. In the empirical component of this paper, we directly examine these two dimensions of variation in a large-scale online shopping experiment, and demonstrate that their quantitative impact on welfare analysis is substantial.

To formalize these arguments, we begin with a model—building on and generalizing Chetty (2009), Chetty, Looney and Kroft (2009, henceforth CLK) and Finkelstein (2009)—of consumers who choose whether or not to purchase a good in the presence of a sales tax. The sales tax is potentially non-salient, and consumers may not correctly account for its presence in their purchasing decisions. Breaking from earlier theoretical treatments of tax salience, we allow for arbitrary heterogeneity in both consumers' valuations for the products and consumers' misreaction to the tax.

We present a series of results that generalize the canonical Harberger (1964) formula for the efficiency costs of taxation. We find that the efficiency cost of imposing a small tax in a previously untaxed market is increasing in the mean of the weights that consumers place on the tax when making purchasing decisions—thus, as in CLK, homogeneous underreaction reduces efficiency costs. However, we additionally show that inefficiency is increasing in the variance of misreactions, to a degree of equal quantitative importance. The result arises because variation in mistakes across consumers generates misallocation of products. When underreaction to the tax is homogeneous, the product is always purchased by those consumers who value it the most, and thus the market preserves the efficient sorting that is obtained with fully optimizing consumers. However, when consumers vary in their misreaction, purchasing decisions depend on both their valuation of the

good *and* on their propensity to ignore the tax, thus breaking the efficient sorting property. The consequences of misallocation are particularly stark when supply is inelastic relative to demand and thus the equilibrium quantity purchased is relatively unaffected by taxation—a situation in which efficiency costs are low when consumers optimize perfectly, but can be substantial in the presence of varying mistakes.

When evaluating "small" taxes, the mean and variance of consumer misreaction—together with the price elasticity of demand—are sufficient statistics for computing efficiency costs. When considering increasing pre-existing taxes, however, accounting for how misreaction changes with the tax rate is crucial. If increases in the tax rate increase attention, and thus "debias" consumers, the distortionary effects of tax increases can be substantially higher than would otherwise be expected. Intuitively, this is because consumers act as if prices have increased not only by the salient portion of the new tax, but also by a portion of the existing tax that they had previously ignored, but now do not.

Taken together, these theoretical results show that empirical estimates of the variation in mistakes are crucial for welfare analysis. However, addressing these questions about variation in mistakes requires datasets containing richer information than simple aggregate demand responses. This motivates our experimental design.

Our experiment studies the behavior of 2,998 consumers—approximately matching the US adult population on household income, gender, and age—drawn from the forty-five US states with positive sales taxes. The experiment utilizes an online pricing task with twenty different non-tax-exempt household products (such as cleaning supplies), and with between- and within-subject variation of three different decision environments. The decision environments induce exogenous variation in the tax applied to purchases, featuring either 1) no sales taxes, 2) standard sales taxes identical to the consumers' city of residence, or 3) high sales taxes that are triple those in the consumers' city of residence. Decisions in the experiment are incentive compatible: study participants use a \$20 budget to potentially buy one of the randomly chosen products, and purchased products are shipped to their homes.

We begin our empirical analysis by estimating the average amount by which study participants underreact to taxes—a statistic, previously estimated in CLK, that is sufficient for welfare analysis when mistakes are homogenous. We find that in the standard tax condition study participants react to the taxes as if they are only 25% of their size. In the triple tax condition, in contrast, study participants react to the taxes as if they are just under 50% of their actual size. Across specifications, this difference in the relative weights that study participants place on taxes in the two conditions is significant at least at the 5% level, and provides initial evidence that consumers are more attentive to higher taxes. Complementing this evidence, we also show that consumers are on average more likely to underreact to taxes on particularly cheap products (posted prices below \$5), than they are to taxes on more expensive products (posted prices above \$5).

Having established variation of misreaction across tax rates, in the second part of our empirical

analysis we focus on variation of misreaction across consumers. This analysis is directly motivated by the efficiency cost formulas that we derive, which show that the efficiency cost of a small tax t on a product sold at price p depends on the variance of underreaction by consumers who are on the margin at p and t. The corresponding statistic of interest is thus the average—computed with respect to the distribution of p and t in the experiment—of $Var[\theta|p,t]$, where θ measures the weight placed on the tax. We bound this statistic through a novel combination of a "self-classifying" survey question and experimental behavior, in a way that requires no assumptions about truth-telling or metacognition. Our estimates of the bound imply that for taxes that are the size of those observed in the US, the variance of consumer mistakes increases the efficiency cost estimate by over 200% relative to what would be inferred under the assumption that consumers are homogeneous in their mistakes.

This paper relates to three distinct literatures. First, beyond extending and generalizing existing the work on tax salience (e.g., CLK, Finkelstein 2009), the paper broadly contributes to a growing theoretical and empirical literature in "behavioral public economics" (see Chetty 2015 for a review, and Mullainathan, Schwartzstein and Congdon 2012 and Farhi and Gabaix 2015 for general theoretical frameworks). Although in our own work we have previously touched on the identification challenges posed by individual differences in mistakes (Allcott and Taubinsky, 2015), and the challenges of relying on locally estimated elasticities when attention is endogenous (Allcott, Mullainathan and Taubinsky, 2014),[1] this paper is the first, to our knowledge, to combine a theoretical framework with an integrated experiment to explicitly bring both facets of variation in mistakes to the forefront, and to directly demonstrate the significant quantitative importance of these issues. These results have immediate applications to the literature on tax misunderstanding;[2] however, our framework for analyzing variation in mistakes is broadly portable, and can serve as a template for empirical analysis of other psychological biases, and in other domains of behavior.

Second, our experimental findings are also relevant to the growing literature on firm and consumer interactions in markets with shrouded attributes (Gabaix and Laibson, 2006; Heidhues et al., 2014; Veiga and Weyl, Forthcoming). The predictions of these models rely on assumptions about heterogeneity, as well as how imperfect processing depends on the size of the shrouded attribute. Our estimates can thus help guide the quantitative predictions of these models.[3]

Third, our work contributes to the literature on boundedly rational value computation (see, e.g., Gabaix, 2014; Woodford, 2012; Caplin and Dean, 2015a; Chetty, 2012). To the best of our knowledge, our result that consumers underreact less to higher tax rates provides one of the first

[1] See also Farhi and Gabaix (2015) for further results relating to these issues, including the importance of attention heterogeneity for Pigouvian taxation, and the implications of misperceptions of and inattention to taxes for income taxation.

[2] For work documenting tax misperceptions see, e.g., Chetty et al. (2013), Chetty and Saez (2013), Bhargava and Manoli (2015) on misunderstanding of the EITC; Abeler and Jäger (2015) for lab experimental evidence about the impacts of complexity; de Bartolome (1995), Liebman and Zeckhauser (2004), Feldman, Katuscak and Kawano (2016) for work related to income tax misperceptions.

[3] Veiga and Weyl (Forthcoming), for example, show that a monopolist's shrouded attribute strategy will depend on the covariance between inattention to the shrouded attribute and household income.

experimental demonstrations in a naturalistic setting of imperfect processing of a financial attribute responding to economic incentives.[4]

The paper proceeds as follows. Section 2 presents our theoretical framework. Section 3 presents our experimental design. Section 4 quantifies average underreaction across different taxes, while Section 5 quantifies the variance of underreaction across consumers. Section 6 utilizes our theoretical framework to discuss the welfare implications of our empirical estimates. Section 7 concludes.

2 Theory

This section analyzes the tax policy implications of variation in consumers' inattention to or misunderstanding of tax instruments. Specifically, we generalize Harberger's (1964) canonical formulas for the efficiency costs of taxation, as well as CLK's formulas for the case of homogeneous consumers. The formulas we develop transparently highlight the importance of accounting for the variation of mistakes across both consumers and tax sizes. The results can be immediately applied to questions about optimal Ramsey or Pigouvian taxes—which we summarize in Section 2.5 and elaborate on in Appendix B—and also apply more broadly to consideration of any kind of imperfectly understood policy instrument. All proofs are contained in Appendix C.

2.1 Set-up

2.1.1 Consumer and Producer Behavior

Consumers: There is a unit mass of consumers who have unit demand for a good x and spend their remaining income on an untaxed composite good y (the numeraire). A person's utility is given by $u(y) + vx$, where $x \in \{0, 1\}$ denotes whether or not the good is purchased, and v is the person's utility from x. Let Z denote the budget (assumed identical across consumers), p the posted price of the product, and t the tax set by the policymaker.[5]

A fully optimizing consumer chooses $x = 1$ if and only if $u(Z - p - t) + v \geq u(Z)$. However, we allow consumers to not process the tax fully. Instead, a consumer chooses $x = 1$ when $u(Z - p - \theta t) + v \geq u(Z)$, where θ—which may covary with v or be endogenous to t—denotes how much

[4]Results on this general topic are mixed. In a lab experiment studying simple vs. complex tax codes, Abeler and Jager (2015) find that study participants underreact to complex changes in the tax code, but that this underreaction does not depend on the magnitude of the change. Hoopes et al. (2015) find that taxpayers pay more attention to capital-gains information when the payoffs to doing so are higher. In tests of boundedly rational decision-making more broadly, Caplin and Dean (2015b) and Caplin and Dean (2013) find that study participants pay more attention to stimuli when given higher incentives, in accordance with a general class of rational inattention models; Allcott (2011) and Allcott (2015) show that consumers pay more attention to energy costs when gasoline prices are higher.

[5]Note that we are assuming here that the policymaker is using a tax instrument with only one level of salience. See Goldin (2015) for a model in which the policymaker can combine tax instruments of differing salience to raise revenue in the least distortionary way possible. Although our analysis could be generalized to consider an optimal mix of more and less salient sales taxes, we suspect that our starting is a reasonable one because of political economy constraints—a politician may have trouble explaining to the public why he chose to break up an otherwise simple tax into shrouded and unshrouded subcomponents.

the consumer under- (or over-) reacts to the tax. Because we make no assumption about the distribution of θ, this modeling approach encompasses a number of psychological biases that may lead consumers to make mistakes in incorporating the sales tax into their decisions. These include:

1. Exogenous inattention to the tax, so that consumers always react to the tax as if it's only θ as big (DellaVigna, 2009; Gabaix and Laibson, 2006).

2. Endogenous inattention to the tax, or boundedly rational processing more broadly, so that consumers pay more attention to higher taxes (Chetty et al., 2007; Gabaix, 2014).

3. Incorrect beliefs, where a person perceives a tax t as $\hat{t}$. In this case, $\theta = t/\hat{t}$.

4. Rounding heuristics.

5. Forgetting about the tax.

6. Any combination of the above biases.

In practice, multiple mechanisms are likely to be in play, and existing data does not shed light on which mechanisms are the most important (CLK), or what the shape of the distribution of θ should look like. Gabaix's (2014) anchoring and adjustment model of attention, for example, predicts that each consumer will have a $\theta \in [0, 1)$, with that value depending on the size of the tax. Other theories of inattention may predict binary attention $\theta \in \{0, 1\}$. Incorrect beliefs and rounding heuristics can generate a variety of different values of θ, with instances in which $\theta > 1$.

We develop our theoretical and empirical framework to be robust to all of these possible mechanisms. Formally, we define θ in terms of consumer behavior. For a given consumer, define $p_{max}(t)$ to be the highest posted price at which the consumer would purchase x at a tax t. Then $\theta := \frac{p_{max}(0) - p_{max}(t)}{t}$. We make no assumptions about the relation between θ and v other than that their joint distribution $F_t(v, \theta)$ generates smooth, downward-sloping demand curves, that $\theta \geq 0$, and that the marginal distribution of v does not depend on t. By allowing the distribution of θ to depend on t we capture the possibility that attention to taxes may depend on the tax rate. With minor abuse of notation, we define $E[\theta|p, t]$ and $Var[\theta|p, t]$ to be the mean and variance of θ of consumers who are indifferent between purchasing the product or not at (p, t).

We let $D(p, t)$ denote demand for x as a function of posted price p and sales tax t. We let D_p and D_t denote partial derivatives with respect to the p and t, and we let $\varepsilon_{D,p}(p, t) = -D_p(p, t)\frac{p+t}{D(p,t)}$ and $\varepsilon_{D,t}(p, t) = -D_t(p, t)\frac{p+t}{D(p,t)}$ denote the elasticities with respect to p and t. We often suppress the arguments p, t in the elasticity to economize on notation.

To focus our analysis on mistakes arising solely from incorrect reactions to the sales tax, we assume that 1) in the absence of taxes, consumers optimize perfectly and 2) consumers' utility depends only on the final consumption bundle (x, y).[6] Welfare analysis under these two assumptions

[6] Assumption 2 implies that we leave out cognitive costs from our efficiency costs and welfare analysis. Although

and our choice-based definition of θ is an application Bernheim and Rangel's (2009) approach to welfare analysis: we view choice in the presence of taxes as provisionally suspect, and we use consumer choice in the absence of taxes as the welfare-relevant frame. We relax the first assumption in Appendix B, following models such as those in Lockwood and Taubinsky (2015) and Farhi and Gabaix (2015).

Producers: We define production identically to CLK: price-taking firms use $c(S)$ units of the numeraire y to produce S units of x. The marginal cost of production is weakly increasing: $c'(S) > 0$ and $c''(S) \geq 0$. The representative firm's *profit* at pretax price p and level of supply S is $pS - c(S)$. Producers optimize perfectly so that the supply function for good x is implicitly defined by the marginal condition $p = c'(S(p))$. Let $\varepsilon_{S,p} = -\frac{\partial S}{\partial p}\frac{p}{S(p)}$ denote the price elasticity of supply. We define $\varepsilon^{TOT}_{D,t} = \frac{d}{dt}D(p,t) \cdot \frac{p+t}{D}$ to be the total percent change in equilibrium demand (taking into account changes in producer prices) caused by a one percent change in the tax

2.1.2 Efficiency Cost of Taxation

We follow Auerbach (1985) in defining the excess burden of a tax for a market with heterogeneous consumers. We let $x_i^*(p,t,Z)$ denote consumer i's choice of $x \in \{0,1\}$ and we let $V_i(p,t,Z) = u(y - px_i^*(p,t,Z) - tx_i^*(p,t,Z)) + v_i x_i^*(p,t,Z)$ denote the consumer's indirect utility function.

We denote the consumer's expenditure function by $e_i(p,t,V)$, which is the minimum wealth necessary to attain utility V under a price p and tax t. Let $R_i(t,Z) = tx_i^*$ denote the revenue collected from this consumer. Excess burden is given by

$$EB = \int_i [Z - e(p_0, 0, V_i(p(t), t, Z)) - R_i(t,Z)] + \pi_0 - \pi_1$$

where $\pi_0 - \pi_1$ is the change in producer profits, p_0 is the equilibrium market price in the absence of taxes, and $p(t)$ is the equilibrium price at tax t. That is, excess burden is the sum of the change in consumer surplus and producer surplus minus government revenue. With quasilinear utility and fixed producer prices (i.e., perfectly elastic supply) this is simply $\int_i (v_i - p_0)(x_i^*(p_0,t) - x_i^*(p_0,0))$: the loss in surplus that accrues from discouraging transactions in which the value of the product v exceeds it's marginal cost of production.

To clarify the key determinants of total excess burden, we write it as a function of two arguments, t and F_t, to clarify its dependence on both the tax and the distribution of θ. The efficiency costs of increasing a tax from t_1 to t_2 can be decomposed into two effects:

$$EB(t_2, F_{t_2}) - EB(t_1, F_{t_1}) = \underbrace{[EB(t_2, F_{t_2}) - EB(t_1, F_{t_2})]}_{\text{Direct distortion effects}} + \underbrace{[EB(t_1, F_{t_2}) - EB(t_1, F_{t_1})]}_{\text{"Nudge channel" distortion effects}} \quad (1)$$

there may be some cognitive costs associated with attention, we do not feel that we have enough evidence to confidently specify a theory of what they should be. Our welfare formulas can be readily extended by including an additional term corresponding to cognitive costs. For small taxes, however, cognitive costs generate a third-order, and thus negligible, efficiency cost. (Chetty et al., 2007).

The first effect corresponds to the direct distortionary effect of the tax, holding the distribution of bias constant. The second effect is the indirect effect that a tax has on excess burden by altering the distribution of consumer bias. The second effect can be understood more broadly as the efficiency costs of a nudge that changes the distribution of consumer bias. To provide a clear exposition of the economics of each of these two effects, we study the two effects in isolation before combining them into one formula.

2.2 Direct Efficiency Costs

For the results presented in the body of the paper, we assume that u is linear (no income effects), but we discuss the implications of income effects at the end of the section, and in more detail in Appendix A.2. To simplify the exposition of some of our main results, we sometimes also assume that θ and v are independently distributed—but we relax that assumption where appropriate in the proofs of the propositions, which provide more general results.

Proposition 1. *Suppose that F_t does not depend on t. Let $p(t)$ denote the equilibrium price as a function of t. Then*

$$\begin{aligned}\frac{d}{dt}EB(t,F_t) &= -E[\theta|p,t]t\frac{d}{dt}D(p(t),t) - Var[\theta|p,t]tD_p(p(t),t) \\ &= E[\theta|p,t]tD(p(t),t)\frac{\varepsilon^{TOT}_{D,t}}{p(t)+t} + \frac{Var[\theta|p,t]}{E[\theta|p,t]}tD(p(t),t)\frac{\varepsilon_{D,t}}{p(t)+t} \qquad (2)\end{aligned}$$

Proposition 1 provides a general formula for the (direct) excess burden of a small tax t when consumers are arbitrarily heterogeneous. When $Var[\theta|p,t]=0$, the formula reduces to the formula provided in CLK, which shows that the excess burden of the tax is proportional to $E(\theta)$. In the simple framework without income effects, the more consumers ignore the tax, the less consumers are discouraged from purchasing the product because of the tax, and thus the smaller the excess burden.

The general formula illustrates that it is not just how much people underreact to the tax on average that matters, but also the variance of (marginal) consumers' underreactions. To take a stark example, suppose that $E(\theta)=0.25$ (for consumers on the margin). When all consumers are homogeneous with $\theta=0.25$, equation (2) shows that the excess burden from a marginal increase in the tax is $(0.25)tD(p,t)\frac{\varepsilon^{TOT}_{D,t}}{p+t}$; that is, the true excess burden is one quarter of what the neoclassical analyst would compute using the tax elasticity of demand. Now, suppose that 25% of the consumers have $\theta=1$ while 75% have $\theta=0$, so that $E[\theta]=.25$ and $Var[\theta]=(.75)(.25)$. In this case, we still have $E(\theta)=0.25$, but (2) implies that the excess burden is now *at least* $tD(p,t)\frac{\varepsilon^{TOT}_{D,t}}{p+t}$, since $\varepsilon_{D,t}\geq\varepsilon^{TOT}_{D,t}$. Interestingly, this is greater than or equal to the inference that would be made by an analyst who assumes that consumers optimize perfectly and thus uses the tax elasticity of demand as a sufficient statistic for calculating excess burden.

The intuition for this result is that heterogeneity in consumers' mistakes creates a market failure that is conceptually distinct from the effect of a homogeneous mistake. If consumers are homogeneous in their underreaction to the tax, then for any quantity of products purchased, the allocation of products to consumers is efficient: the product is still purchased by consumers who derive the most value from it. When consumers are heterogeneous in their underreaction, however, there is misallocation: the consumers purchasing the product are now not just the consumers who derive the most value from it, but also consumers who underreact to taxes the most. There is thus an additional efficiency cost from an inefficient match between consumers and products.[7]

Another important insight from Proposition 1 is that the efficiency costs arising from misallocation depend on the elasticity of the demand curve, rather than on the elasticity of the *equilibrium* quantity of x in the market. Thus measurement of (changes of) the equilibrium quantity is not sufficient to calculate efficiency costs, even when combined with estimates of average underreaction—this is in stark contrast to standard efficiency cost of taxation results, as well as Chetty (2009)'s results that allow endogenous producer prices but assume homogeneous underreaction. This is most clear in the case of inelastic supply:

Corollary 1. *Suppose that supply is inelastic* ($\varepsilon_{S,p} = 0$) *and that* F_t *does not depend on* t*. Then*

$$\frac{d}{dt}EB = \frac{Var(\theta|p,t)}{E(\theta|p,t)} tD(p(t),t)\frac{\varepsilon_{D,t}}{p(t)+t}$$

Corollary 1 shows that when supply is inelastic—and thus the equilibrium quantity produced by the market does not change—the excess burden of a small tax t depends only on the variance of bias and the price elasticity of demand. Intuitively, this is because all of the efficiency cost is generated by misallocation, the extent of which is proportional to the variance of θ—which quantifies the extent of individual differences—and the price elasticity of demand—which determines how much the individual differences translate to different purchase decisions. That efficiency costs can be significant even when supply is inelastic is in sharp contrast to standard results in public finance that efficiency costs should be zero if taxes do not distort the equilibrium quantity. More generally, the results imply that when consumers are heterogeneous in their underreaction, efficiency costs will be significantly higher than in the standard model when supply is relatively inelastic compared to demand.[8]

[7] This point about misallocation and departure from traditional deadweight loss analysis can be obtained in some neoclassical settings as well. Glaeser and Luttmer (2003) show that rent control not only distorts the equilibrium quantity purchased, but also creates an allocational failure whereby properties are no longer purchased by the consumers who value them the most.

[8] Empirical work on how the supply elasticities compare to demand elasticities is scarce and has not settled on a range. Studies that estimate pass-through of salient consumption taxes (those included in the upfront price of the good) find that the pass-through to the final, after-tax price—given by $\frac{\varepsilon_{S,p}}{\varepsilon_{S,p}-\varepsilon_{D,p}}$—ranges from 19% to 48% (Benzarti and Carloni, 2016). Studies that estimate pass-through of not-fully-salient sales taxes into the after-tax price—given by $\frac{\varepsilon_{S,p}-(1-E[\theta|p,t])\varepsilon_{D,p}}{\varepsilon_{S,p}-\varepsilon_{D,p}}$—find estimates ranging from 70% to 100% (Besley and Rosen, 1999; Doyle and Samphantharak, 2008).

The formula in Proposition 1 can also be used to extend the classic Harberger (1964) second-order approximations of the efficiency costs of taxation. We begin by quantifying the efficiency costs of introducing a small tax t into a previously untaxed market. Although Proposition 1 characterizes only direct efficiency costs, it can be used to provide a *complete* characterization of the excess burden of introducing a small tax t in a previously untaxed market. Because the nudge channel distortion effect is irrelevant when there are no pre-existing taxes (as per equation 1, $EB(0, F_t) - EB(0, F_0) = 0$), in this case the only relevant efficiency costs are the direct efficiency costs. We thus have:

Proposition 2. *If terms of order t^3 or higher are negligible, the excess burden is*

$$EB(t, F_t) \approx \frac{1}{2} t^2 D \left[E[\theta|p,t] \frac{\varepsilon_{D,t}^{TOT}}{p(t)+t} + Var[\theta|p,t] \frac{\varepsilon_{D,p}}{p(t)+t} \right]$$

The nudge distortion channel is not irrelevant when there are pre-existing taxes, but we now use Proposition 1 to characterize the *direct* efficiency costs of increasing pre-existing taxes. We maintain the standard assumptions of the "Harberger Trapezoid" formula (Harberger, 1964) that for $k \geq 2$, the terms $t(\Delta t)^k D_{pp}$, $t(\Delta t)^k S_{pp}$, $(\Delta t)^{k+1}$ are negligible. This assumption corresponds to cases in which the demand and supply curves are approximately linear, or to cases in which both the pre-existing tax t and the change Δt are sufficiently small (or a suitable combination of the two).

Proposition 3. *Suppose that $F_{t_1} = F_{t_2} \equiv F$ for $t_2 = t_1 + \Delta t$. Then, if for $k \geq 2$ the terms $t(\Delta t)^k D_{pp}$, $t(\Delta t)^k S_{pp}$, $(\Delta t)^{k+1}$ are negligible, and if θ and v are distributed independently, the excess burden of increasing the tax from t_1 to t_2 is*

$$\begin{aligned} EB(t_2, F) - EB(t_1, F) &\approx - \left(t_1 \Delta t + \frac{(\Delta t)^2}{2} \right) \left(E[\theta|p(t_1), t_1] \frac{d}{dt} D(p(t), t)|_{t=t_1} + Var[\theta|p(t_1), t_1] D_p(p(t_1), t_1) \right) \\ &= \left(t_1 \Delta t + \frac{(\Delta t)^2}{2} \right) D(p(t_1), t_1) \left(\frac{E[\theta|p(t_1), t_1] \varepsilon_{D,t}^{TOT}}{p(t_1)+t_1} + \frac{E[\theta|p(t_1), t_1] \varepsilon_{D,p}}{p(t_1)+t_1} \right) \end{aligned}$$

Like Proposition 1, Proposition 3 shows that the standard formula is modified in two ways. First, the change in the equilibrium quantity, $\frac{d}{dt} D(p(t), t)|_{t=t_1}$, is now multiplied by the average θ of marginal consumers. Second, increasing taxes increases misallocation of products to consumers, which leads to a new term given by the product of the variance of θ and the price elasticity of demand.

2.3 Indirect Efficiency Costs: The Consequences of Debiasing

In this section, we provide "Harberger-type" formulas for the efficiency costs (or benefits) of changing the distribution of θ. We keep the tax fixed, and we consider a family of distributions $F_n(\theta, v)$ that are smooth functions of n for all θ, v. We think of n as the "nudge parameter," and we ask

how the excess burden of a tax changes as we shift this parameter by some small amount from n to $n + \Delta n$. The formulas here serve as an intermediate step to the final formulas that we derive in Section 2.4, but we also view them to be of independent interest as a novel extension of the standard public finance toolbox. We provide results under two different conditions:

Condition A For all n and θ, $F_n(\theta + h(n)) = F_0(\theta)$, for some differentiable function $h(n)$.

Condition B The term $t^{k+1}\frac{\partial^k}{\partial p^k}D$ is negligible for $k \geq 2$, and $\theta \perp v$ for all n.

Condition A is that the nudge translates the distribution uniformly by some level $h(n)$. The first part of condition B is a variation of the standard Harberger formula assumption that the term $t(\Delta t)^k D_{pp}$ is negligible, but is a slightly stronger requirement on how small t or D_{pp}, needs to be. The second part of condition B is simply that θ and v are distributed independently.

To appreciate the need for placing additional structure on the distributions, consider the difficulty of generally estimating efficiency costs in the seemingly simple case in which θ takes on just two possible values, θ_1 and θ_2, and is distributed independently of v. Let $EB_i(t)$ denote the excess burden arising from the type θ_i consumers. The efficiency cost of increasing the measure of type θ_2 consumers by some small amount dn is then $(EB_2(t) - EB_1(t))dn$. But if t is not small and the demand curve of each θ is highly nonlinear so that each θ type's *price* elasticity varies with i, we have no way of quantifying $EB_2(t) - EB_1(t)$ in terms of observables. Further structure is needed to relate the demand curves of the different θ types in terms of observables.

The additional structure provided by Condition A is that it ensures that the nudge essentially behaves like a tax, affecting all consumers uniformly. The additional structure provided by Condition B is that essentially ensures a quadratic approximation for the efficiency costs corresponding to each θ type, and that the price elasticities of demand are not too different across the θ types.

For the results in this section, we let D^{F_n} denote the demand curve under F_n and E_{F_n} the expectation operator with respect to F_n. To simplify exposition, we will also assume that producer prices are fixed, but we present general formulas with endogenous producer prices in the proofs in the appendix .

Proposition 4. *Suppose that producer prices are fixed ($\varepsilon_{S,p} = \infty$), and that either Condition A or Condition B is satisfied. Then*

1. $\frac{d}{dn}EB(t, F_n) \approx -\frac{d}{dn}\left(E_{F_n}[\theta^2|p,t]\right)t^2 D_p^{F_n}$.
2. *If for $k \geq 3$ the terms $(\Delta n)^k$ are negligible then*

$$EB(t, F_{n+\Delta n}) - EB(t, F_n) \approx -\frac{1}{2}t^2\left(E_{F_{n+\Delta n}}[\theta^2|p,t] - E_{F_n}[\theta^2|p,t]\right)D_p^{F_n}$$

The intuition behind Proposition 4 is straightforward. As we have already established, efficiency costs depend on both the mean and the variance of θ. Consequently, the welfare impacts of a nudge should correspond to how the nudge impacts the mean and variance of θ. This is exactly the result of Proposition 4, as $E[\theta^2|p,t] = E[\theta|p,t]^2 + Var[\theta|p,t]$.

2.4 Total Efficiency Costs

We now combine our results from Sections 2.2 and 2.3 to quantify the total efficiency costs of taxation. As in Section 2.3, we focus on fixed producer prices to simplify exposition.

Proposition 5. *Consider two taxes t_1 and $t_2 = t_1 + \Delta t$. Suppose that producer prices are fixed ($\varepsilon_{S,p} = \infty$), that θ is distributed independently of v, and that either Condition A or Condition B is satisfied for the family of distributions F_t indexed by the tax t. Suppose also that for $k \geq 2$, the terms $t(\Delta t)^k D_{pp}$ and $(\Delta t)^k$ are negligible. Then*

$$\begin{aligned} EB(t_2, F) - EB(t_1, F) &\approx -\left(t_1(\Delta t) + \frac{(\Delta t)^2}{2}\right)\left(E[\theta|p, t_2]^2 + Var[\theta|p, t_2]\right) D_p \quad (3) \\ &- \frac{1}{2}t^2\left(E[\theta^2|p, t_2] - E[\theta^2|p, t_1]\right) D_p \quad (4) \end{aligned}$$

Proposition 5 is essentially a combination our earlier results about the direct efficiency costs of a tax and our results about the efficiency costs of a nudge. Line (3) corresponds to the direct efficiency costs (as in Proposition 3)[9], while (4) corresponds to the nudge channel efficiency costs (as in Proposition 4).

The formula in Proposition 5 is written in its most compact form using the *price* elasticity of demand. One might be tempted to think that using tax elasticities could eliminate additional terms corresponding to costs of debiasing, since the tax elasticity captures both the direct and indirect effects that increasing a tax has on demand. However, simply using tax-elasticity version of the direct efficiency costs formula in Proposition 3 will still not account for all of the efficiency costs, because it is not just the change in demand that matters, but also how the values v of the marginal types change. We clarify in the corollary below.

Corollary 2. *Under the assumptions of Proposition 5, and the assumption that the approximations $E[\theta|p, t_2] - E[\theta|p, t_1] \approx \Delta t \frac{d}{dt} E[\theta|p, t]|_{t=t_1}$ and $Var[\theta|p, t_2] - Var[\theta|p, t_1] \approx \Delta t \frac{d}{dt} Var[\theta|p, t]|_{t=t_1}$ are valid, efficiency costs can also be expressed as*

$$\begin{aligned} EB(t_2, F) - EB(t_1, F) &\approx \frac{\left(t_1(\Delta t) + \frac{(\Delta t)^2}{2}\right) D}{p + t_1}\left(\frac{E[\theta|p, t_1] + E[\theta|p, t_2]}{2}\varepsilon_{D,t} + \frac{Var[\theta|p, t_1] + Var[\theta|p, t_2]}{2}\varepsilon_{D,p}\right) \\ &+ \frac{1}{2}t(\Delta t + t)\frac{D}{p + t_1}\left(Var[\theta|p, t_2] - Var[\theta|p, t_1]\right)\varepsilon_{D,p} \\ &+ \frac{t(\Delta t)}{2}\frac{D}{p + t_1}\left(E[\theta^2|p, t_2] - E[\theta^2|p, t_1]\right)\varepsilon_{D,p} \end{aligned}$$

To illustrate the formula in the corollary, suppose that θ is homogenous, so that $Var[\theta|p, t] = 0$. In this case, efficiency costs are not simply given by $\left(t_1(\Delta t) + (\Delta t)^2/2\right)\frac{D}{p+t_1}E[\theta|p, t_1]\varepsilon_{D,t}$, as would be prescribed by Proposition 3. There are additional efficiency costs, arising from the nudge

[9] Note that under the assumptions of Proposition 3, $E[\theta|p, t_2]^2 D_p = E[\theta|p, t_2] D_t = E[\theta|p, t_1] D_t$.

effect, given by $\frac{t(\Delta t)}{2}\frac{D}{p+t_1}\left(E[\theta|p,t_2]^2 - E[\theta|p,t_1]^2\right)\varepsilon_{D,p}$. In the simple case of $Var[\theta|p,t] = 0$, these additional efficiency costs correspond to the fact that the value of the product to the marginal consumer under t_2 is not simply $p + E[\theta|p,t_1](t_1 + \Delta t)$, as it would be if taxes did not change underreaction, but is instead $p + E[\theta|p,t_2](t_1 + \Delta t)$. That is, in contrast to the standard model, the value of the product to the marginal consumer is a convex, rather than a linear function of the tax when $E[\theta|p,t]$ is increasing in t.

2.5 Extensions and Optimal Tax Implications

Optimal Ramsey and Pigouvian Taxes The formulas we present for quantifying how changes in the tax affect welfare or excess burden have direct implications for optimal taxes. In Appendix B we derive optimal tax formulas in a Ramsey framework, using a more general model that allows for other market frictions arising from either externalities or other imperfections in consumer choice (i.e., the possibility that consumers misoptimize even in the absence of taxes, or that they spend their remaining income suboptimally on the composite untaxed good).

In formalizing the implications of our excess burden calculations for optimal taxes, the results in the appendix generate several new insights. First, when there are no other market frictions and taxes are used only to meet a fixed revenue requirement, the optimal tax system may deviate from the canonical Ramsey inverse elasticity rule in several ways. If people underreact less to taxes on more expensive products, that implies that other things equal, the tax rates on bigger ticket items should be smaller. Holding product prices constant, the inverse elasticity rule is also dampened if θ is on average increasing in the tax. This is because increasing taxes increases deadweight loss through the additional debiasing channel.

Second, we characterize how taxes depend on other market imperfections, and consider whether a less salient tax is optimal for the policymaker, building on the analysis in Farhi and Gabaix (2015). When there is no variation in θ, underreaction to the tax is always beneficial, even in the presence of externalities (or internalities). Because the consumers who buy the product are still those who value it the most, any not-fully-salient tax can still be set high enough to achieve the socially optimal consumption of x. With variation in θ, however, the more salient tax is better if the externality is sufficiently large relative to the value of public funds. This is because introducing a not-fully-salient tax causes misallocation, and thus cannot achieve the socially optimal consumption of x. Our general message about the importance of taking into account the misallocation arising from heterogeneity in θ is thus particularly relevant in the presence of other market frictions.

Income Effects We have thus far assumed that $u(y)$ is linear, so that there are no income effects. This is a reasonable assumption for small ticket items for which p and t are small relative to income. The analysis of income effects is more complicated, but follows the same principles as the baseline excess burden formula without income effects, and thus we relegate this to Appendix A.2. In the appendix, we formally show that the formulas we derive in the body of the paper still hold

in the presence of income effects when either 1) the taxed product is a small share of consumers' expenditures or 2) the taxed product is purchased on a reasonably frequent basis, and the consumer can observe his budget in between the purchases. Thus for common household commodities, we believe that our results hold robustly in the presence of income effects.

However, for infrequent, large-ticket purchases there can still be efficiency costs when consumers ignore the tax fully. This can occur when a consumer spends more money than he realizes on the product in question, and thus consumes inefficiently too little y in the future after he is surprised with how small his budget is.

Distributional concerns In Appendix A.3 we also extend our framework to incorporate distributional concerns. We show that with redistributive concerns, the relative regressivity costs of not fully salient sales taxes, as compared to fully salient sales taxes, are determined by how the the mistakes—given by $(\theta_i - 1)^2$ and reflecting either under or over-reaction to the tax—covary across the income distribution.

2.6 Identification from Aggregate Demand Data

What kinds of datasets identify the statistics necessary for welfare analysis? CLK and Chetty (2009) show that for a representative consumer, the generalized demand curve $D(p,t)$ identifies excess burden when pre-existing taxes are small. Under these assumptions, θ is identified by the average degree of underreaction to taxes relative to prices, $D_t(p,t)/D_p(p,t)$.

In Appendix A.1 we prove two main results about identification of efficiency costs under more general assumptions. First, we focus on the case in which $F(\theta|p,t)$ is degenerate for all p,t, and show that when θ is endogenous to the tax rate, locally-estimated elasticities no longer identify θ or excess burden, although full knowledge of $D(p,t)$ does. Intuitively, this is because the ratio of demand responses D_t/D_p is roughly equal to $E[\theta|p,t] + \frac{d}{dt}E[\theta|p,t]t$, and thus identifies $E[\theta|p,t]$ only when the distribution of θ does not depend on t. Thus datasets containing only local variation in t are not sufficient for questions about the efficiency costs of non-negligible increases in sales taxes.

Second, we show that if θ can be heterogeneous, conditional on p and t, then $D(p,t)$ can never identify the dispersion, and thus welfare. While the average θ is identified by D_t/D_p for small taxes, the variance of θ is left completely unidentified. These results show that key questions about the variation of underreaction to taxes cannot be identified from existing data sources. This motivates our experimental design.

3 Experimental Design

Platform The experiment was implemented through *ClearVoice Research*, a market research firm that maintains a large and demographically diverse panel of participants over the age of 18. This platform is frequently used by firms who ship products to consumers to elicit product ratings, but

is additionally available to researchers for academic use (for other examples of research using this platform, see, e.g., Benjamin et al. 2014; Rees-Jones and Taubinsky 2016). Two key features of this platform make it appropriate for our experimental design. First, ClearVoice provides samples that match the US population on basic demographic characteristics. Second, ClearVoice maintains an infrastructure for easily shipping products to consumers, which facilitates an incentive-compatible online-shopping experiment.

Overview Figure 1 provides a synopsis of the experimental design. The design had four parts: 1) elicitation of residential information, 2) module 1 shopping decisions, 3) module 2 shopping decisions, and 4) end-of-study survey questions. The design is both within-subject—we vary tax rates for a given consumer between modules 1 and 2—and between-subject—consumers face different tax rates in module 1. Decisions are incentivized: study participants have a chance to receive a $20 shopping budget to actually enact their purchasing decisions, and ClearVoice ships any products purchased. The between-subject aspect of the design allows us to test for anchoring or demand effects, while the within-subject aspect of the design increases statistical power and provides identification that is not possible from between-subject aggregate data.

Each consumer was randomly assigned to one of three arms: 1) the "No Tax Arm," 2) the "Standard Tax Arm," and 3) the "Triple Tax Arm." The purpose of the no tax arm was to identify any order effects on valuations over the course of the experiment and to help test for demand or anchoring effects.[10]

Each module consisted of a series of shopping decisions involving 20 common household products. In module 1, consumers made shopping decisions with either a zero tax rate (no tax arm), a standard tax rate corresponding to their city of residence (standard tax arm), or a tax rate equal to triple their standard tax rate (triple tax arm). In module 2, consumers in all three arms made decisions in the absence of any sales taxes. The same 20 products were used in each module and in each arm of the experiment. The order in which the 20 products were presented was randomly determined, and independent between the two modules.

Because our experimental design involves language about the sales tax rate that study participants pay in their city of residence, to avoid confusion we asked ClearVoice to only recruit panel members from states that have a positive sales tax. This excluded panel members from Alaska, Montana, Delaware, New Hampshire, and Oregon. The remaining 45 states are all represented in our final sample. Prior to learning the details of the experiment, consumers were asked to state their state, county, and city of residence. To correctly determine the money spent in the experiment, this information was matched to a dataset of tax rates in all cities in the United States.[11]

[10] An additional goal was to identify the distribution of random shocks to valuations between module 1 and module 2, and to combine this with the other two arms to deconvolve the distribution of individual θ parameters from the distribution of measurement error. Ultimately, the variance of the measurement error we encountered was too high to permit a well-powered deconvolution of this type.

[11] Local tax rate data is drawn from the April, 2015 update of the "zip2tax" tax calculator.

This design is closely related to several recent experimental studies of tax salience (e.g., Feldman and Ruffle 2015; Feldman et al. 2015), but differs in important ways. Our design combines within-subject manipulation of tax rates with a pricing mechanism that elicits full and precise demand curves. This design, combined with our unusually large sample size, allows us to infer the sufficient statistics of our general welfare formulas—an exercise not possible with previous experimental designs.

Purchasing Decisions Each product appeared on a separate screen. For each product, consumers saw a picture and a product description drawn from Amazon.com. Consumers then used a slider to select the highest *tag price* at which they would be willing to purchase the product. It was explained that "The tag price is the price that you would find posted on an item as you walk down the aisle of the store; this is different from the final amount that you would pay when you check out at the register, which would be the tag price, plus any relevant sales taxes." Figure 2 shows examples of the decision screen.

If a study participant selected the highest price on the slider, \$15, he was directed to an additional screen where he was asked a hypothetical free-response question about the highest tag price at which they would be willing to buy the product.

The three different decision environments were described to consumers as follows:

- *No tax decision environment:* In the *no tax* decisions consumers were told that "In contrast to what shopping is like at your local store, no sales tax will be added to the tag price at which you purchase a product." It was explained that "You can imagine this to be like the case if there were no sales tax, or if sales tax were already included in the prices posted at a store." As depicted in Figure 1, the no tax decisions constituted the second module that consumers encountered in each experimental arm, and also the first module that consumers encountered in the no tax arm.

- *Standard tax decision environment:* For the *standard tax* decisions the instructions prior to decisions were that "The sales tax in this section of the study is the same as the standard sales tax that you pay (for standard nonexempt items) in your city of residence, [city], [state]." The standard tax decisions constituted the first module of the standard tax arm.

- *Triple tax decision environment:* For the *triple tax* decisions the instructions prior to decisions were that "The sales tax in this section of the study is equal to triple the standard sales tax that you pay (for standard nonexempt items) in your city of residence, [city], [state]." The triple tax decisions constituted the first module that consumers encountered in the triple tax arm.

To make this experimental shopping experience as close as possible to the normal shopping experience, and to enable tests for incorrect beliefs, consumers were not told what tax rate applies in

their city of residence. Once consumers read the instructions (and answered the comprehension questions), they were never reminded of the taxes again in the tax modules. In contrast, the no tax modules emphasized the absence of taxes to ensure that choices in those models reflect consumers' true willingness to pay for the produces. Example decision screens are shown in Figure 2.

Incentive Compatibility Decisions in the experiment were incentive compatible. All study participants who passed the necessary comprehension questions (described below) had a 1/3 chance of being selected to receive a $20 budget; accounting for the probability of failing the comprehension check, this chance was approximately 1/4. Participants were informed of this incentive structure prior to making any decisions, but they did not know if they received the budget until they completed the experiment. If they did not receive the budget, they simply received a compensation of $1.50 and no products from the study. Consumers who were selected to receive the $20 budget had one of their decisions implemented. To avoid confounds arising from income effects, only one out of the forty decisions (from modules 1 and 2 combined) was selected to be played out. Outcomes were determined using the Becker-DeGroot-Marshak (BDM) mechanism. A random tag price was drawn between 0 and 15, and if it was below the maximum tag price the consumer was willing to pay, then the product was sold to the consumer. In the event that the product was sold to the consumer at tag price p, a final amount of $p(1+\tau)$ (where τ is the experimentally induced tax rate) was subtracted from this consumer's budget, and the product was shipped to the consumer by ClearVoice. Participants received a full explanation of the BDM mechanism, and were also told that it was in their best interest to always be honest about the highest tag price at which they would buy the product.

Comprehension Questions As we will discuss in Section 4.7, it is important to ensure that study participants read the instructions explaining what tax rate applies to their decisions, so that our results are not confounded by subjects simply ignoring or misreading the instructions. In both module 1 and module 2, we thus gave study participants a multiple choice comprehension question asking them about the final amount they would pay if they purchased a product at a particular tag price.In both modules, the quiz question appeared on the same screen as the instructions for that module.

Product Selection To arrive at the final list of 20 household products, we began with a list of 75 potential items in the $0 to $15 price range compiled by a research assistant. From this list, we eliminated items that were tax exempt in at least one state. We then ran a pre-test with ClearVoice to elicit (hypothetical) willingness to pay for the items. We selected 20 items that had unimodal distributions of valuations and had the least censoring at $0 and $15. Appendix F lists the products, prices, and Amazon.com product descriptions that were displayed to study participants.

Survey Questions After completing the main part of the experiment, study participants received a short set of questions eliciting household income, marital status, financial literacy, ability to compute taxes, and health habits. We discuss these questions in further detail in the analysis.

ClearVoice also collects and shares various demographic information on its panel members, including educational attainment, occupation, age, sex, and ethnicity. We report these basic demographics in Section 4.1.

4 Quantifying Underreaction Across Different Tax Sizes

4.1 Sample Selection, Demographics, and Balance

4.1.1 Sample

A total of 4,328 consumers completed the experiment. For this sample, 3,066 correctly answered instruction-comprehension questions in both module 1 and module 2. For our primary analysis, we exclude those who did not answer both comprehension questions correctly. Out of the remaining 3066 consumers, 30 consumers were not willing to buy at a positive price in at least one of their decisions. Because our primary estimates are formed using the logarithm of the ratio of module 1 and module 2 prices, we cannot use at least one observation for each of these 30 consumers. We thus exclude them from analysis as well. We additionally exclude 10 consumers who reported living in a state with no sales tax.[12]

In part due to our pretest for product selection, only 0.9% of all responses were censored at \$15. For responses that were censored, we use consumers' uncensored responses to the hypothetical question about the maximum tag price. However, this question did not force a response, and 28 consumers did not provide an answer to this question upon encountering it. We exclude these consumers as well, leaving us with a final sample size of 2,998.

Unsurprisingly, the 29% of consumers who did not pass the comprehension questions do not react to the differences in taxes across conditions. We exclude these consumers from our main sample, because the misoptimization these consumers exhibit is likely due to misunderstanding of our experimental manipulation. This type of misunderstanding is conceptually distinct from misunderstanding a given tax rate, and not the object of interest in our theoretical analysis. While we exclude these subjects as a matter of principle, in Appendix E.3 we show that all of our key results are robust to the using the full sample.[13]

[12] These 10 consumers were erroneously recruited for the study because they had recently changed residence and that information was not yet updated in ClearVoice's records.

[13] We also included questions to check if participants understood the BDM mechanism. 78% of participants passed those comprehension questions, and we show in Appendix E.5 that our results are robust to restricting to this sample. We are far less concerned about potential misunderstanding of the pricing mechanism for two reasons. First, participants were clearly instructed that it was in their best interest to always truthfully report the maximum tag price at which they would be willing to buy the product. Second, most forms of "strategic" price reporting do not confound estimates of θ (only estimates of true valuations): even if participants scale their price thresholds up or down because of systematic misperceptions of the mechanism, the ratio of the bids will still be $1+\theta\tau$ (in expectation).

4.1.2 Demographics and Balance

Table 1 presents a summary of the demographics of our final sample. All participants in the final sample are over the age of 18, and all but thirty-one participants are over the age of 21. Our final sample—which is 49% male, has a median income of $50,000, average age of 50, and is 41% college-educated—is similar to the US population on these basic demographics. We fail to reject the null hypotheses of equality of the demographics in table 1 when comparing Arm 1 vs. Arm 2 (F-test $p = 0.49$), Arm 2 vs. Arm 3 (F-test $p = 0.94$) or Arm 1 vs. Arm 3 (F-test $p = 0.36$).

There are, however, small but statistically significant differences in the likelihood that consumers pass the comprehension questions in the different arms of the study.[14] A possible reason is that consumers in the no tax arm answer the same comprehension question twice because the decision environment is identical in both modules. Consumers in the tax arms, however, answer two different comprehension questions, and are thus more likely to incorrectly answer at least one. Moreover, consumers in the triple tax arm might be especially likely to incorrectly answer the quiz question if they simply assume that any sales tax would be the standard one they face in their city of residence. We will show in Section 4.7, that our results are robust to both worst-case assumptions about differential selection and the inclusion of the full sample.

4.2 Summary of Behavior

We begin with a graphical summary of the data. Figure 3 provides a summary of the demand curves as functions of before-tax price. To construct the figure, we start with demand curves $D_k^{\mathrm{C},m}(p)$ for each product k, where C $\in$\{0x, 1x, 3x\} denotes the experimental arm, m denotes the module, and p the before-tax price. Because there are 20 products, we summarize the data by plotting the *average demand curves* $D_{avg}^{\mathrm{C},m}(p) := \frac{1}{20}\sum_k D_k^{\mathrm{C},m}(p)$ for each arm C and module m. The demand curves have a "zig-zag" pattern because a significant portion of consumers choose tag prices that are near whole dollar amounts.

Panel (a) of Figure 3 shows that consumers do react to sales taxes in module 1, as their willingness to buy at a given before-tax price is decreasing in the size of the sales tax. But while consumers in the different arms behave differently in module 1, panel (b) of Figure 3 shows that there is no evidence of anchoring or demand effects in module 2, where all consumers face the same no tax environment. The panel shows that the average demand curves are nearly identical in module 2, which is confirmed by several statistical tests. For our first test, we compute an average pre-tax price $\bar{p}^i = \frac{1}{20}\sum_k p^{ik}$ for each consumer i, and then compare the distributions of $\bar{p}^i$. Kolmogorov-Smirnov tests find no differences in the $\bar{p}_i$ between the no tax and standard tax arms $p = 0.73$), between the no tax and triple tax arms ($p = 0.29$), and between the standard tax and triple tax arms $p = 0.50$). In Appendix E.1, we also show OLS and quantile regressions comparing

[14] The likelihood of correctly answering both comprehension questions are 78%, 70%, and 65% in the no-tax, standard-tax, and triple-tax arms, respectively. The null hypothesis of equal pass rates is rejected for any pair of arms at the 5% significance level.

the average willingness to pay in module 2, which similarly detect no differences.

While consumers react to taxes, they do not react to taxes as much as perfect optimization would require, as we show in Figure 4. Panel (a) of Figure 4 is identical to panel (a) of Figure 3, while panel (b) shows average demand as a function of total, tax-inclusive price. If consumers reacted to the taxes fully, the demand curves in panel (b) of Figure 4 would be identical. However, the figure shows that consumers do not react to taxes fully, and are willing to buy at higher final prices in the presence of taxes, particularly large ones.

4.3 Econometric Framework

We now present our baseline econometric framework for studying how underreaction to taxes varies by experimental condition and by observable characteristics. Let p_1^{ik} be the highest tag price a subject i is willing to pay in module 1 for product k, and define p_2^{ik} analogously for module 2. Note that in the absence of noise or order effects, $p_2^{ik}/p_1^{ik} = 1 + \theta_{ik}\tau_i$, where $1 - \theta_{ik}$ is the degree of underreaction to the tax on product k by consumer i. Thus for a consumer i in either the standard or triple tax arms, $\frac{y_{ik}}{\tau_i} \approx \theta_{ik}$, where τ_i is the tax rate faced by the consumer in module 1 and $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$.

Of course, $\frac{y_{ik}}{\tau_i}$ may be a noisy estimate of θ_{ik} because study participants' reported values for the product fluctuate, and it may also be biased if perceived values of the products vary between phase 1 and phase 2 (what we call "order effects"). We allow order effects to depend on any estimated covariates, but we assume that order effects do not vary by experimental arm in our final sample. This assumption, labeled A1 below for reference, allows us to identify order effects from behavior in the "no tax" treatment arm.

A1 For any vector of covariates X_{ik}, $E[y_{ik} - \log(1 + \theta_{ik}\tau_i)|X_{ik}]$ does not depend on τ_i.

For a vector of covariates X_{ik} we will estimate the following model:

$$\begin{aligned} E[y_{ik}|X] &= \log(1 + \theta_{ik}\tau_i) + \beta X_{ik} \\ E[\theta_{ik}|X_{ik}] \approx E\left[\frac{\log(1 + \theta_{ik}\tau_i)}{\tau_i}|X_{ik}\right] &= \alpha X_{ik} \end{aligned}$$

The model above implies the following moment conditions:

$$E[X_{ik}' y_{ik}] = X_{ik}'\beta X_{ik} \quad \text{for no tax arm} \tag{5}$$

$$E\left[X_{ik}'\left(\frac{y_{ik} - \beta X_{ik}}{\tau_i}\right)\right] = X_{ik}'\alpha X_{ik} \quad \text{for std./triple tax arms} \tag{6}$$

Equation (5) identifies any order effects in the data using the no tax arm. These order effects are partialed out from y_{ik} in the standard and triple tax arms in equation (6), which allows us to estimate $E[\theta_{ik}]$ as a linear function of covariates X_{ik}. When estimating (5) and (6) for either the standard or triple tax arm separately, the system of equations is exactly identified. When pooling

data from multiple treatment arms, we will assume that (6) holds independently for each arm, but with a common α. The system is thus over-identified, and we use the two-step GMM estimator to obtain an approximation to the efficient weighting matrix. We will often condition on $p_2^{ik} \geq \underline{p}$ (typically $p_2^{ik} \geq 1$)—i.e., focusing analysis on those with non-negligible willingness to pay—as a means of increasing precision. Because most of our analysis takes p_2/p_1 as an object of interest, noisiness in responses can generate dramatic variation in this quantity when valuations approach zero. All of our point estimates are robust to the inclusion of all data.

Although our approach may seem somewhat complicated, we show in Appendix E.4 that all of our main results are robust to a simpler approach,.using OLS to regress y_{ik} on the tax rate. As we elaborate in that appendix, however, we prefer our GMM approach as it avoids the need to assume that mean underreaction is constant across tax sizes within an experimental arm–an assumption our results refute.[15]

4.4 Average Underreaction to Taxes by Experimental Arm

Table 2 presents our estimates of average θ in each arm using the econometric framework presented in Section 4.3. We provide estimates using all data, as well after conditioning on $p_2^{ik} \geq 1$ and $p_2^{ik} \geq 5$ (as in Section 4.3, p_m^{ik} denotes the highest before-tax price at which a consumer i will buy product k in module m). The table shows that across all specifications, we estimate an average θ of just over 0.25 in the standard tax arm and an average θ of just under 0.5 in the triple tax arm. The standard errors on these estimates are sufficiently tight to reject both that consumers completely neglect taxes and to reject that consumers react to the taxes fully. All the estimates are more precise in the second and third columns than in the first column, as the ratio p_2^{ik}/p_1^{ik} is naturally most noisy when a consumer attaches low value to the product. We will thus continue conditioning on $p_2^{ik} \geq 1$ throughout the rest of our analysis.

The difference in average θ between the arms is significant at the 5% level when using all data or when conditioning on $p_2^{ik} \geq 1$, and it is significant at the 0.1% level when conditioning on $p_2^{ik} \geq 5$.[16]

[15]Note, also, that in principle, we could have used $1 - p_2^{ik}/(\tau_i p_1^{ik})$ instead of y_{ik} as the dependent variable. We don't do this because using the raw ratio p_2^{ik}/p_1^{ik} gives more weight to outliers, and thus the estimates are unduly influenced by the inclusion or exclusion of the top 1% of values of p_2^{ik}/p_1^{ik}. Because of this extreme right tail of the distribution of p_2^{ik}/p_1^{ik}, a strategy for decreasing the weight on extreme realizations is necessary to stabilize the estimates. Estimates in our preferred specification using the log transformation are very similar to the estimates that are obtained after winsorizing at least the top 1% of values of $1 - p_2^{ik}/(\tau_i p_1^{ik})$ for each arm.

[16]Feldman, Goldin and Homonoff (2015, henceforth FGH) run a complementary lab experiment with 227 Princeton students to study purchasing behavior at a 8% vs. a 22% sales tax rate, similar to our standard vs. triple tax conditions. The three arms of the FGH experiment are similar in structure to ours, although there are important differences that prevent direct comparability. While the FGH experiment was not designed to identify average θ by experimental condition (or by covariates), the statistic that the FGH design does allow estimation of is $\frac{1-E[\theta|8\%]}{1-E[\theta|22\%]}$, where $E[\theta|x\%]$ is the average θ in the condition with an $x\%$ tax rate. This statistic is estimated to be 0.4 with a standard error of 0.75, and a 95% confidence interval of [0,1.86]. By comparison, we estimate $\frac{1-E[\theta|\text{standard}]}{1-E[\theta|\text{triple}]}$ to be 1.42 with a standard error of 0.175 and a 95% confidence interval of $[1.08, 1.77]$. Thus, while our 95% confidence interval is nested within the FGH 95% confidence interval, the significantly greater power of our design allows us to reject the null hypothesis that the ratio equals 1—the necessary threshold for establishing an increase in attention.

4.5 Further Tests of Endogenous Attention

Our baseline results suggest that consumers attend more to higher taxes, though there are several caveats. First, consumers might overreact to the triple tax if they are surprised by the unusual scenario (Bordalo, Gennaioli, and Shleifer 2015). Second, our estimates of average θ in the triple tax arm may be biased downward because it may take time for people to develop new heuristics for how they respond to the larger taxes.

A complementary analysis that addresses these caveats would be to estimate whether consumers attend more to taxes in states with larger sales taxes. However, such an analysis would require a sample size that is approximately 45 times larger than ours to be well-powered, and our data has nothing meaningful to say about this question.[17] We can, however, ask a different complementary question about whether consumers attend more to taxes on higher priced products, since the total tax $t = \tau p$ is increasing in the posted price p. We operationalize this by dividing all consumers (from all three arms) into three bins corresponding to their module 2 valuation ($p_2^{ik} < 5$, $p_2^{ik} \in [5, 10)$, and $p_2^{ik} \geq 10$), and then estimating an average θ for each of the three bins.

Columns (1)-(3) of Table 3 report the results of this estimation. Column 1 presents estimates for the standard tax arm; column 2 presents estimates for the triple tax arm; and column 3 presents estimates for the pooled data. When pooling data, we allow for different baselines of average θ for the different arms but, to maximize power, we assume that the impact of moving to a higher bin is the same across the arms. Although we are underpowered for this analysis in the standard tax arm, the table shows that when pooling the data, or when restricting to the triple tax arm, consumers in the second and third bin have a higher average θ than consumers in the first bin. The differences in average θ are approximately 0.12 for second vs. first bin in the pooled analysis and 0.15 for third vs. first bin in the triple tax arm or pooled analysis. We do not detect a difference (although we also cannot reject a moderate one) for average θ between the second and third bin—this suggests that attention may not increase linearly with price and that consumers employ different attention strategies for very low price products below \$5 vs. moderate price products above \$5.

This analysis is consistent with average θ increasing in the absolute tax $p\tau$. However, this result could also be obtained if consumers who are willing to pay the most for the products are also the consumers who have the highest θ. Columns (4)-(6) report an analogous test, ruling out this possibility through the inclusion of individual fixed effects (Appendix D.2 formally documents how we modify our GMM strategy). While estimates are slightly attenuated toward zero as compared to the estimates in the first three columns, we again see the similar pattern: on average, θ appears to be somewhat smaller when $p_2^{ik} < 5$ than when $p_2^{ik} \geq 5$.

[17] The average tax rate of the bottom 50% of tax rates is 6.4%, while the average tax rate of the top 50% tax rates is 8.3%. Thus the difference in average θ between the top and bottom quantiles should be only $(8.3/6.4-1)/(3-1) = 0.15$ as big as the difference in average θ between the standard and triple tax arms, assuming that average θ scales linearly with the size of the tax rate. To estimate this effect with the same level of precision that we estimate the difference between the standard and triple tax arms, we would thus need a sample size that is $(1/0.15)^2 \approx 44.4$ times as large.

4.6 Sources and Correlates of Consumer Mistakes

4.6.1 Do Consumers Know the Tax Rates?

To assess consumers' knowledge of the sales tax rates, and whether underestimation of the tax rates generates some of the underreaction, we included the following survey question at the end of the study: "What percent is the sales tax rate in your city of residence, [city], [state]? If your city exempts some goods from the full sales tax, please indicate the rate for a standard nonexempt good. If you're not sure, please make your best guess."[18]

Although the question asked participants to enter their answer as a percent, a small minority of participants appears to not have read the instructions and entered their answer as a decimal (e.g. 0.07 instead of 7%). For the 6% of participants who entered an answer below 0.1, we assume that they did not enter their answer as a percent, and thus we convert their answer by multiplying it by 100.

On average, consumers' beliefs are very accurate. 52% of consumers know their tax rate exactly, about 74% are within 0.5 percentage points, and about 85% are within 1 percentage point. The average of beliefs is 7.05%[19], while the average actual tax rate of consumers in the study is 7.32%, indicating almost no mean bias.

To provide a graphical summary of how perceived beliefs vary with the actual tax rate, we construct Figure 5 which plots average perceived taxes for each of 25 quantiles of actual taxes. The best-fit regression line in the figure has an intercept of 0.27 percentage points (s.e. 0.32) which is not statistically different from 0 ($p = 0.41$) and a slope of 0.93 (s.e. 0.04) which is marginally statistically different from 1 ($p = 0.09$).[20] Thus, although there is some weak evidence that beliefs are not completely unbiased, the possible bias is still very small in magnitude and cannot account for the vast underreaction to taxes that our subjects exhibit. We conclude that incorrect beliefs are a negligible source of consumer mistakes, consistent with CLK's survey results from consumers in a California store.

4.6.2 Demographic Covariates

In Appendix E.2, we analyze how average θ varies by demographic covariates, including income, financial literacy, ability to compute taxes, age, sex, marital status, education, and race. When pooling data across both arms, we find that demographics have significant explanatory power (F-test $p < 0.01$). We find a significant positive association between average θ and financial literacy, and a marginally significant positive association between average θ and income and numeracy. And we find a statistically significant negative association between θ and age. We find no relationship between θ and sex, marital status, education, and race.

[18] If a small minority of participants misreported their city of residence, then our results are a lower bound on how well participants know their actual sales tax rate.

[19] For this statistic we exclude four obvious outlier values that are above 100.

[20] For both Figure 5 and the regression estimate, we also exclude the four outlier values that are above 100.

Of these results, perhaps the most economically significant result is that average θ is marginally significantly higher for consumers in the fourth quartile of the income distribution than for consumers in the first quartile of the income distribution. To the extent that the propensity of mistakes varies by income groups, the presence of non-salient taxes will impact the regressivity of sales taxes—a point previously explored in Goldin and Homonoff (2013), and which we formalize in our heterogeneous model in Appendix A.3.

4.7 Robustness to Selection on Comprehension Questions

A limitation of any experiment other than a natural field experiment is the possibility that the experiment confuses subjects in a manner that natural environments do not. In our context, we were concerned that even fully optimizing subjects might misunderstand our assignment of tax rates to experimental conditions, and thus create the appearance of misunderstanding of tax rates. For this reason, our final sample includes only study participants who correctly identified the experimentally induced tax incomprehension questions before both module 1 and module 2.

Selecting on comprehension, however, may introduce differences in the final samples across the three arms. While our samples do not detectably differ on observable demographic characteristics, we cannot rule out that they differ on unobservable characteristics that are correlated with θ. In fact, there are slight but statistically significant differences in the likelihood of correctly answering the comprehension questions across treatment arms. However, in Appendix D.1 that we can obtain a tight lower bound for the difference between average θ in the triple and standard tax arms under mild assumptions. When implementing the lower bound, we find that we can reject no difference between the triple and standard tax conditions at the 10% significance level ($p = 0.08$) when using all data, at the 5% significance level ($p = 0.04$) when conditioning on module 2 price $p_2^{ik} \geq 1$, and at the 1% significance level ($p < 0.01$) when conditioning on $p_2^{ik} \geq 5$. Thus, while we were concerned *ex ante* about the possibility of selection induced by our screening criteria, *ex post* it appears that our results are robust to this concern.

5 Quantifying the Variation of Underreaction Across Consumers

Having established that underreaction varies across both economic and demographic conditions, we now return to our first theoretical question: quantifying individual differences in θ.

As the results in Section 2 show, the statistic needed for welfare analysis is $Var[\theta|p, \tau]$—the variance of θ for consumers who are indifferent between buying the product or not at posted price p and tax rate τ. The statistic we aim to estimate is thus $E_{p_1,\tau}[Var[\theta|p_1, \tau]]$; that is, our variance of interest averaged over all (p_1, τ) pairs. Note that $E_{p_1,\tau}[Var[\theta|p_1, \tau]] \leq Var[\theta]$, and that this inequality is strict if θ varies with τ and p_1. Consequently, simply estimating the variance of θ would produce upward-biased estimates of how much variance is coming from individual differences, because this statistic would also include variation in θ due to differences in p_1 and τ.

Informally, the idea behind our approach is to to partition study participants into subgroups with different average θ's based on self classifications. When then compute the variance of the subgroup means, which provides a lower bound for the total variance. We divide subjects into subgroups using our "self-classifying" survey question, which we *ex ante* selected as most promising to be predictive of underreaction, and which indeed turned out to be our most predictive measure *ex post*. In this section, we begin by presenting the details of our self-classifying survey question. We the present our methodology in Section 5.2 and implement an estimate of the lower bound in Section 5.3

5.1 The Self-Classifying Survey Question

The self-classifying survey question asked consumers in the standard and triple tax arms the following: "Think back to Section 1, where you made your first twenty decisions about tag prices. In that section, there was a sales tax that you would have to pay if you bought an item from that section. If there was no sales tax in Section 1, would you choose higher tag prices for the products?" The possible answers to the question were "Yes," which we code as $R = H$; "Maybe a little," which we code as $R = M$; and "No," which we code as $R = L$. Table 4 summarizes participants' responses to the survey question. Overall, approximately 10% of participants answered "Yes," approximately 55% answered "Maybe a little," and approximately 35% answered "No." Participants in the triple tax arms were more likely to say "Yes" or "Maybe" than participants in the standard tax arm (Ranksum test $p < 0.01$).[21]

Responses to this question are highly predictive of experimental behavior. To estimate an average θ for each survey response, we employ the same methodology as in Section 4.3, with the exception that because this survey question was not asked in the no tax arm, we make the additional assumption A2 that if survey responses R are predictive of behavior, it is solely because they are correlated with θ:

A2 $E[y_{ik}|\theta_{ik}, R] = E[y_{ik}|\theta_{ik}]$

A2 implies that for the standard and triple tax arms,

$$E\left[\frac{y_{ik} - E[y_{ik}|\text{no tax arm}]}{\tau_i}|R = r\right] = E\left[\frac{\log(1+\theta_{ik}\tau_i)}{\tau_i}|R = r\right], \tag{7}$$

Thus $E\left[\frac{\log(1+\theta_{ik}\tau_i)}{\tau_i}|R = r\right]$ can now be estimated as in Section 4.3.

Table 5 shows that this survey question has a striking degree of predictive power. The table shows that roughly, the average θ is not statistically different from 0 for consumers who answer "No," is in the neighborhood of 0.5 for consumers who answer "Maybe a little", and is in a neighborhood

[21] However, the difference is not large in magnitude, despite being statistically significant. One possible reason for the minor difference is "relative thinking" ((Bushong et al., 2015): because taxes were much larger in the triple tax arm, what participants in the triple tax arm considered a large response to the tax was likely different than what participants in the standard tax arm considered a large response to the tax.

of 1 for consumers who answer "Yes." Table 5 thus shows that under assumption A2, there are stark differences in θ between different consumers. Moreover, the remarkable predictive power of the survey question suggests that, consistent with models of bounded rationality and deliberate attention, people are aware of the mistakes they make in responding to sales taxes.

However, these results do not yet prove that there are individual differences *conditional on a price-tax pair* (p_1, τ). Given our results about how the distribution of θ covaries with the tax size, it is possible that some of these differences may be driven by variation in θ across the pairs (p_1, τ). To quantify individual differences *conditional on a price-tax pair* (p_1, τ), we proceed with the development of our lower-bound estimator.

5.2 A Lower-Bound for the Variance of Mistakes: Theory

Let R be the random variable of study participants' responses to the survey question, which can take on the values $R = H$, $R = M$ or $R = L$.[22] We now create new random variables $\phi := \frac{\log(1+\theta\tau)}{\tau}$, $\mu := E[\phi|p, \tau]$, $\bar{\phi} := E[\phi|R = r, p, \tau]$. In words, ϕ is the approximation to θ that we obtain from our log-transformed data. The variable μ is the average of ϕ for all consumers who are marginal at price p and tax rate τ. And the variable $\bar{\phi}$ takes on three different values for consumers marginal at price p and tax rate τ: amongst the marginal consumers with $R = r$, it is the average of ϕ for those consumers. For short-hand, we set $\bar{\theta}_r := E[\bar{\phi}|R = r]$; this is the average ϕ across all consumers with $R = r$ (without conditioning on a price-tax pair).

Proposition 6.

$$
\begin{aligned}
E_{p_1,\tau}[Var[\theta|p_1, \tau]] &\geq E\left[Var[\bar{\phi}|\tau, p_1]\right] && (8)\\
&\geq Pr(R = H)\left(E[\bar{\phi}|R = H] - E(\mu|R = H)\right)^2 && (9)\\
&+ Pr(R = M)\left(E[\bar{\phi}|R = M] - E(\mu|R = M)\right)^2 && (10)\\
&+ Pr(R = L)\left(E[\bar{\phi}|R = L] - E(\mu|R = L)\right)^2 && (11)
\end{aligned}
$$

Proposition 6 shows that $E_{p_1,\tau}[Var[\theta|p_1, \tau]]$ can be bounded from below by the significantly easier to estimate expression in (9)-(11). The expression in (9)-(11) is similar to $Var[\theta_R]$; that is, to the variance of the three-point distribution that puts mass $Pr(R = H)$ on $\bar{\theta}_H$, mass $Pr(R = M)$ on $\bar{\theta}_M$, and the remaining mass on $\bar{\theta}_L$. The difference is that the conditional means $E[\mu|R]$ are not necessarily equal to the mean of the three-point distribution, which is the unconditional mean $E[\mu] = E[\theta]$. By using the conditional means $E[\mu|R]$ in each term in (9)-(11), the expression corrects for the fact that $Var[\bar{\theta}_R]$ would overestimate $E_{p_1,\tau}[Var[\theta|p_1, \tau]]$ if all individual differences in θ were due simply to variation in (p_1, τ).

[22] Our technique can be immediately generalized to any observable characteristic R that can take on any number of finite values.

In words, the conditional mean $E[\mu|R]$ is constructed as follows: 1) compute the average $\phi \approx \theta$ for each pair (p_1, τ), which is μ, and then 2) compute the average μ with respect to the (induced) conditional distribution of (p_1, τ) given $R = r$. As an example, suppose that $R = H$ was associated only with value $p_1 \geq 10$, $R = M$ was only associated with values $p_1 \in [5, 10)$, and $R = L$ was only associated with values $p_1 < 5$. This corresponds to a case in which all variation in survey answers is captured by variation in p_1. In this case, we would have that $E[\mu|R = r] = \bar{\theta}_r$ for each r, and thus the lower bound in (9)-(11) would be zero.

The idea behind the proof, which is contained in the appendix, is as follows. First, we show that $E_{p_1,\tau}[Var[\theta|p_1, \tau]] \geq E\left[Var[\frac{\log(1+\theta\tau)}{\tau}|\tau, p_1]\right]$, which follows because the concave log transformation is a contraction and thus reduces variance. Second, we use the fact that conditional on each (p_1, τ), the distribution of ϕ is a mean-preserving spread of the distribution of $\bar{\phi}$. This establishes $Var[\phi|p_1, \tau] \geq Var[\bar{\phi}|p_1, \tau]$ for each (p_1, τ), and thus that $E_{p_1,\tau}[Var[\theta|p_1, \tau]] \geq E\left[Var[\bar{\phi}|p_1, \tau]\right]$. Third, we arrive at the final quantity in (9)-(11) through an application of the Cauchy-Schwarz inequality.

Although in principle one could attempt to use self-classifying survey questions to estimate the statistic in (8), in practice it is a highly multidimensional statistic that is estimable to a far lower degree of precision than the statistic in (9)-(11).[23]

5.3 A Lower Bound for the Variance of Mistakes: Estimation

A challenge in estimating the lower bound from Proposition 6 is estimating the terms $E(\mu|R = r)$. Because our dataset is finite, we cannot obtain an estimate of each $\mu(p_1, \tau)$ for each pair (p_1, τ). Instead, we partition the price-tax space into small cells of positive measure, and estimate an average value of $\frac{\log(1+\theta_{ik}\tau_i)}{\tau_i}$ within each cell. Formally, let $\{\boldsymbol{p_j}\}_{j=1}^{15}$ denote the fifteen cells $[0, 1], [1, 2], \ldots, [14, \infty)$ and let $\{\boldsymbol{\tau}_j\}_{j=1}^{5}$ denote the five cells $(0, 6\%], [6\%, 7\%], \ldots [9\%, \infty)$. Because only 0.5% of all prices are above \$15, and only 0.1% of all taxes are above 10%, we simply include these observations in the last cells without much loss of precision. Denote by $\mathbf{p}(p)$ the cell containing p, and denote by $\boldsymbol{\tau}(\tau)$ the cell containing τ. We approximate $\mu(p_1, \tau)$ by

$$\tilde{\mu}(p_1, \tau) = E\left[\frac{\log(1 + \theta_{ik}\tau_i)}{\tau_i}|p_1^{ik} \in \mathbf{p}(p_1), \tau_i \in \boldsymbol{\tau}(\tau)\right]. \tag{12}$$

As the cell sizes converge to zero, $\tilde{\mu}$ will converge to μ. To estimate the lower bound we simply replace each theoretical moment with it's empirical moment counterpart, and we bootstrap the standard errors of the estimators. See Appendix D.3 for details of the empirical implementation.

Table 6 presents the results. The top row displays our estimates of (42) for both the standard

[23] Estimating (8) would involve the average of many squares of terms, with each term measured with noise. In contrast, the bound in (9)-(11) first collapses the first moments from (8) into only three averages, and then takes the squares of those averages. Thus the bound in (9)-(11) can be estimated much more precisely for the same reason that the variance of an average of n random variables is smaller than the average of the variance of those n random variables.

and triple tax arms. The point estimates are 0.133 for the standard tax arm, and 0.094 for the triple tax arm. To benchmark these estimates, consider what the variances would be if all consumers either processed the tax fully ($\theta = 1$) or completely neglected it ($\theta = 0$). Given a mean of 0.25 in the standard tax arm, the variance would then be $0.25 - 0.25^2 = 0.19$ in that arm. Given a mean of approximately 0.5 in the triple tax arm, the variance would be $0.5 - 0.5^2 = 0.25$ in that arm. Thus our lower bound estimates are approximately 70% and 37% of what the variances would be in the perfectly binary cases of the single- and triple-tax arms, respectively.

To compute standard errors and the mean bias of our estimator, we use the percentile block bootstrap (with 1000 iterations), sampling at the consumer level. As the second row shows, there is a small mean bias of approximately 0.01 for the standard tax arm, implying that a bias-corrected estimate is 0.124.[24] The bias is nine times smaller in the triple tax arm because all effect sizes are three times larger, and thus the relative variance of noise is nine times smaller.

We compute approximate 95% confidence intervals in two ways: 1) using the standard percentile method, and 2) using the (median-) bias-corrected percentile method. As with mean bias, the median bias is reassuringly small, and thus both methods produce similar approximations to the 95% confidence intervals. Importantly, we find that even the 5% confidence bounds are large enough to substantially increase the efficiency costs of taxation, as we show in Section 6.1.

5.4 Alternative Approaches

In this section, we discuss the advantages of our bounding approach relative to two alternative implementations.

As a first alternative, notice that our experimental design allows us to calculate an estimate of θ for each experimental subject, since each of the 20 within-subject product evaluations provides a noisy estimate of this parameter. Examining the distribution of these estimates provides a seemingly simple, but heavily confounded, way of inferring the distribution of θ. The variance of the distribution of individually estimated coefficients reflects both by the true variance in θ—our object of interest—as well as the approximation error inherent in making a small-sample inference—a confounding term. Implementing this strategy in our data does suggest an enormous degree of variance; however, most of this estimate is driven by the noisiness of the estimates.

This approach could, in theory, be modified to deconvolve the variance of measurement error (i.e., random fluctuation in BDM valuations) and the variance of misreaction. Indeed, the no tax condition of our experiment was designed to identify the variance of the measurement error term, so long as two concerns were avoided. As a practical concern, if the variance of measurement error encountered in this arm is large enough, a deconvolution approach would be ill-powered. As a theoretical concern, the presence of rounding heuristics in BDM responses (i.e., rounding

[24]The source of the bias is that any noise in our estimates of $\bar{\theta}_r$ or $E[\mu|R = r]$ amplifies the statistic in (42) because it involves squares of noisily estimated moments. For example, even if the true value of the statistic in (1) was zero, our estimates would still be positive simply because of noisiness in y_{ik}.

one's valuation to the nearest whole number) generates additional variance that confounds the deconvolution; though this issue does not confound first-moment estimates and thus our bounding approach. Since our data exhibits both substantial measurement error and strong evidence of rounding, we implement the bounding approach described above.[25]

While these alternative strategies do have the benefit of providing a point estimate of the variance of misreaction, we believe the practical and technical considerations favor the use of our more robust and conservative bounding approach.[26]

6 From Empirical Magnitudes to Welfare Implications

We now use the theoretical results from Section 2 to translate the experimental results from Sections 4 and 5 into the welfare estimates that those empirical magnitudes would imply. We assess our welfare estimates relative to a benchmark that assumes that misreaction is exogenous and homogeneous, and find that this benchmark substantially understates the welfare costs of taxation.

6.1 Individual Differences

To translate the estimates from Section 5.3 into excess burden estimates, we use the formula in Proposition 2, which expresses excess burden in terms of the mean and variance of θ. To provide maximally conservative estimates, we suppose that supply is perfectly elastic, because as shown in Proposition 2, the relative importance of individual differences increases as the elasticity of supply decreases.[27] For the illustrative calculations here, we set $E[\theta|p,t]$ to equal our estimate of average θ, and we bound $Var[\theta|p,t]$ with our lower-bound estimate of $E_{p,t}[Var[\theta|p,t]]$.

Let $EB_{neoclasical}$ denote the excess burden that would be calculated by a neoclassical analyst who assumes that consumers are not biased, and who relies on the elasticity of demand with respect to the tax.[28] Let $EB_{homogeneous}$ be the excess burden that would be computed by an analyst who assumes that θ is homogeneous, and knows the mean θ from, say, estimating D_t/D_p.[29] Finally, let EB denote the actual excess burden.

Consider now the implications of heterogeneity for welfare inferences. For the standard tax arm, $EB_{homogeneous} \approx (0.25)EB_{neoclassical}$. However, by Proposition 2, the actual excess burden is $EB \geq (0.25 + 0.124/0.25)EB_{neoclassical} = (0.75)EB_{neoclassical}$. For the triple tax arm,

[25] In practice, about 40% of the decisions in our study are close within a few cents of a round number, suggesting that subjects' do engage in some rounding behavior.

[26] A final, and more technical advantage of our approach is that we do not need to assume that the second or higher order moments of the distribution of noise are identical across the arms.

[27] And as discussed in Section 2.5 and further in Appendix A.2, income effects exacerbate excess burden, with that additional effect also increasing in the variance of the bias.

[28] That is, $EB_{neoclassical} = \frac{1}{2}t^2 D(p,t)\frac{\varepsilon_{D,t}}{p+t}$

[29] As shown CLK (and replicated in Proposition 7 for unit demand), the ratio D_t/D_p identifies θ for homogeneous consumers for small t. As shown in the proof of Proposition 8 and implicitly used in the result, it is more generally true that $D_t/D_p = E[\theta|p,t]$ for small t.

$EB_{homogeneous} \approx (0.48)EB_{neoclassical}$. However, by Proposition 2, the actual excess burden is $EB \geq (0.48 + 0.094/0.48)EB_{neoclassica} = (0.68)EB_{neoclassical}$.

Thus for the standard tax arm, individual differences inflate excess burden by over 200% compared to a representative agent calculation, and actually bring the overall estimates closer to the neoclassical case. For the triple tax arm, individual differences inflate excess burden by over 40% as compared to a representative agent calculation. We stress that these estimates are lower bounds—both because we use lower bounds for the variance of θ, and because we assume supply is perfectly elastic—and that the actual impact of individual differences is likely to be much greater.

6.2 Endogenous Attention

We now turn to the implications of endogenous attention that we formalize in Proposition 5. For the calibration, we take $\Delta t = 2t$, and we set $E[\theta|t] = 0.25$ and $E[\theta|t+\Delta t] = 0.5$, consistent with the experimental results. To maintain the same benchmark and units throughout the whole section, we again compute the impact of endogenous attention against the benchmark of homogeneous and exogenous θ. Under the assumption that $F(\theta|p,t)$ is degenerate, Proposition 5 implies that

$$EB(t+\Delta t) - EB(t) \approx \left(t\Delta t(0.5)^2 + \frac{(\Delta t)^2}{2}(0.25)^2 \right) D_p + \frac{t^2}{2}(0.5^2 - 0.25^2) \approx 1.09t^2$$

Consider now inferences under the assumption of homogeneous and exogenous θ. Suppose that the analyst computes $E[\theta] = 0.25$ by studying responses to standard taxes. Then assuming exogenous (and homogeneous) θ, the analyst would infer the excess burden of tripling the tax to be $4t^2(.25)^2 = 0.25t^2$. In this case, the endogeneity of θ with respect to t implies that the correct estimate is more than 400% higher.[30]

7 Discussion

In this paper, we have shown that in addition to measuring the "average mistake," measuring the variation in mistakes is crucial for questions about policy design. When there are individual differences in underreaction to a not-fully-salient sales tax, this increases the efficiency costs arising from that tax's distortionary effect on demand. When underreaction varies with economic incentives, this affects the demand response to new policies and introduces a new channel by which taxes distort behavior. Estimates from our experimental population suggest that these dimensions of variation exist, are sizable in magnitude, and can starkly affect the welfare analysis of tax policies.

These issues are of course not unique to sales taxes, and arise in any question about tax policy. And more broadly, these issues arise in any setting where the true price of a good is divided into different components of differing salience. The theoretical framework we develop in Section 2 can be

[30]The analysis above could be repeated to take variance of θ into account by substituting our lower bound variance estimates—this yields very similar result.

easily extended to accommodate related shrouded attributes, and can therefore serve as a template for robust behavioral welfare analysis.

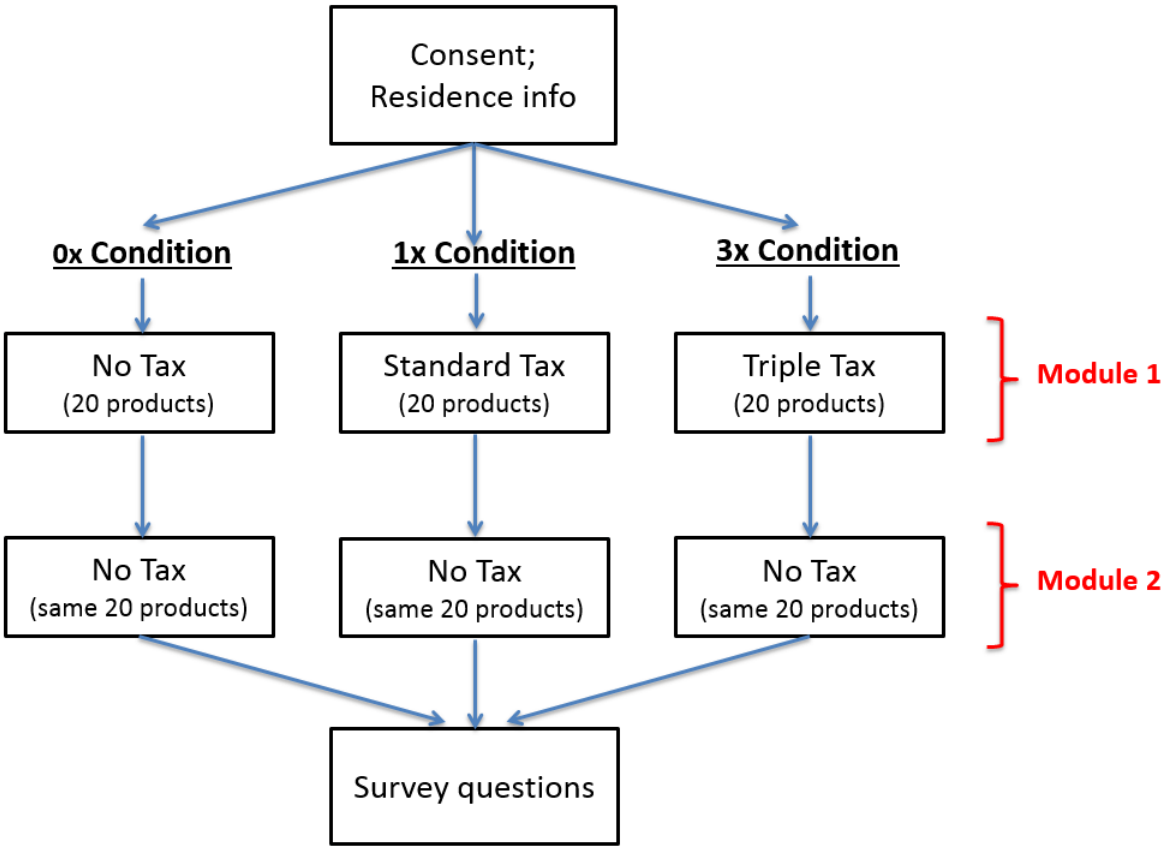
While we believe our theoretical framework is broadly portable, caution is needed when using our experimental estimates to assess welfare in external settings. When implementing our experiment, we devoted significant effort and resources to recruiting a broad and diverse subject population, and to making our experiment as natural as possible while still providing the necessary within-subject measurements. However, as with any experiment, important external-validity concerns remain. We discuss our two main concerns below.

First, we emphasize that our experiment relied on the use of the Becker-DeGroot-Marshak procedure to measure willingness to pay. While useful for precise, incentive-compatible elicitations of demand curves, we worry that this mechanism could trigger a different psychology than simply deciding whether or not to purchase a given item.

Second, the population used in our study is likely non-representative. Despite matching the US population on several key observable demographics, unobserved characteristics could influence selection into our online survey platform. However, were heterogeneity in mistakes not present in the general population, it would not be found in arbitrary subsamples; as such, we do not view these issues as a hindrance to a demonstration that meaningful heterogeneity exists. We view our measurement of these statistics as an initial step, and proof of concept, of a necessary empirical agenda working toward robustly incorporating heterogeneity into behavioral welfare analysis.

As this agenda progresses, it will both benefit from, and inform, the explicit modeling of the psychology of bounded-rationality. In principal, refined and vetted models of attention would place useful structure on our forecasts of heterogeneity in mistakes, and thus the corresponding implications for welfare. We aim to pursue the refinement of these models, and their integration into welfare analysis, in future work.

Figure 1: Experimental Design



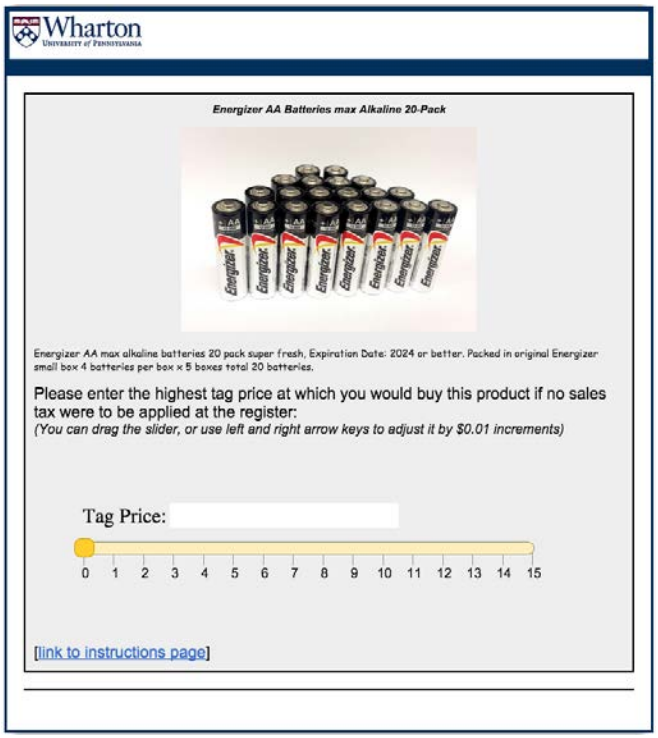
Notes: This figure summarizes our experimental design. For full details, see the accompanying discussion in Section 3.

Figure 2: Decision Format

(a) Tax Module



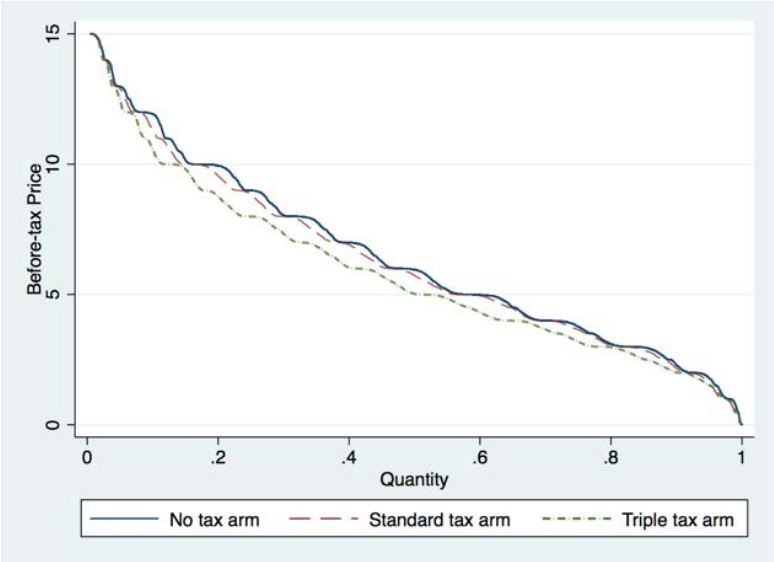
(b) No Tax Module



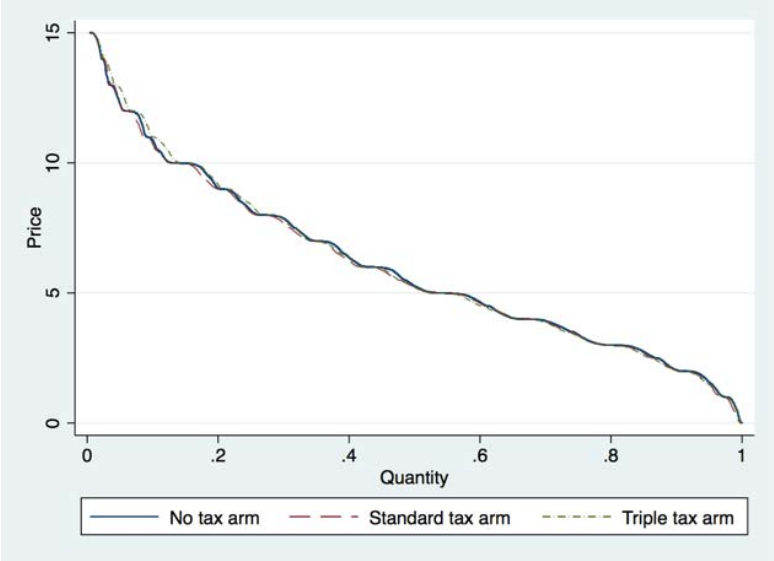
Notes: Panel a shows an example of a pricing decision from modules where taxes apply. Consumers indicate the highest tag price at which they would buy the product. As in typical shopping environments—and as was explained in the experimental instructions—the final price that applies at "check out" is the tag price plus sales taxes. Panel b shows an example of a pricing decision from modules where taxes do not apply. As can be seen in the prompt, respondents are instructed to consider the case where no sales tax is added at the register.

Figure 3: Average Demand Curves in the First and Second Stages of the Experiment

(a) Demand as a function of before-tax price in the first stage of the experiment, where consumers face different tax rates



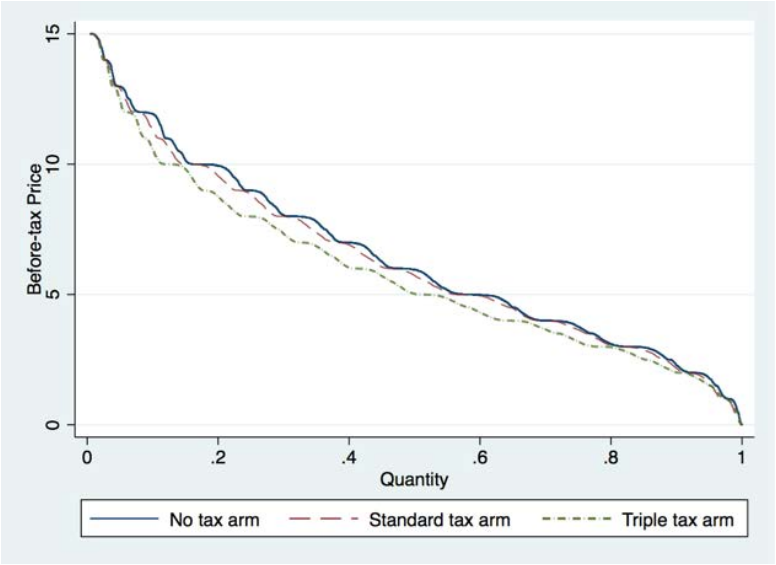
(b) Demand in the second stage, where there are no additional sales taxes in any arm



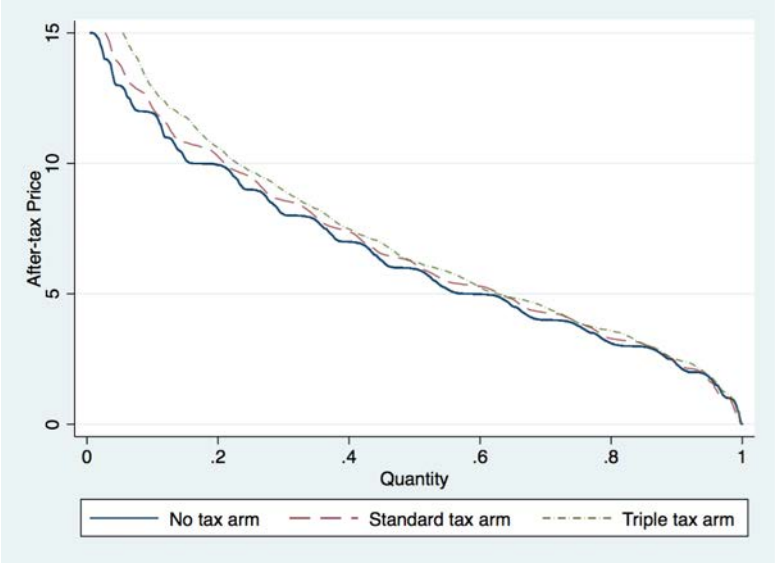
Notes: This figure plots demand curves from the first and second modules of the experiment, averaging across all 20 products.. In the first stage, consumers face either no (additional) taxes, standard taxes, or triple their standard taxes. In the second stage, consumers in all three arms face no additional taxes. For the first stage, we plot demand curves as functions of the before-tax prices. To construct the figure, we start with demand curves $D_k^{\mathrm{C},m}(p)$ for each product k, where C $\in\{$0x, 1x, 3x$\}$ denotes the no-tax, standard-tax or triple-tax experimental arm, m denotes the module (stage), and p the before-tax price. The average demand curves are $D_{avg}^{\mathrm{C},m}(p) := \frac{1}{20}\sum_k D_k^{\mathrm{C},m}(p)$.

Figure 4: Average Demand Curves in the First Stage of the Experiment, as Functions of Before- vs. After-Tax Price

(a) Demand as a function of before-tax price in the first stage of the experiment

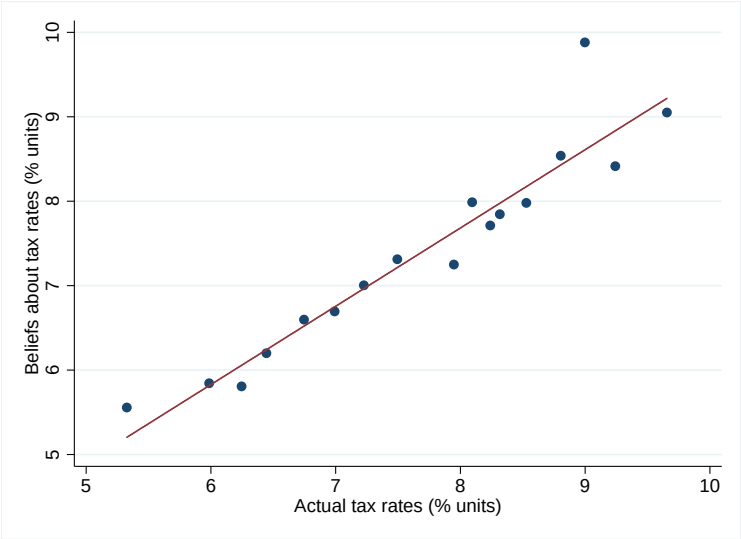


(b) Demand as a function of after-tax price in the first stage of the experiment



Notes: This figure plots the average of the demand curves for the 20 products in the first stage of the experiment. In the first stage, consumers face either no (additional) taxes, standard taxes, or triple the standard taxes. In the second stage, consumers in all three arms face no additional taxes. We plot demand curves as functions of the before-tax prices in panel (a), and as a functions of the after-tax prices in panel (b). To construct the figure, we start with demand curves $D_k^{\mathrm{C},m}(p)$ for each product k, where C $\in$\{0x, 1x, 3x\} denotes the no-tax, standard-tax or triple-tax experimental arm, m denotes the module, and p the before- or after-tax price. The average demand curves are $D_{avg}^{\mathrm{C},m}(p) := \frac{1}{20}\sum_k D_k^{\mathrm{C},m}(p)$.

Figure 5: Perceived vs. Actual Sales Tax Rates



Notes: This figure plots the relationship between the actual tax rates subjects face and the tax rates they believe apply. To construct the figure, we first divide the actual tax rates into 25 quantiles. We then plot the average belief vs. the average actual tax rate for each of the quantiles.

Table 1: Demographics by Experimental Arm

	All	No Tax	Std. Tax	Triple tax	F-test p-val
Age	50.49	50.80	50.43	50.20	$p = 0.66$
	(14.63)	(14.27)	(14.83)	(14.82)	
Household Income	63.04	61.86	63.67	63.78	$p = 0.68$
	(56.29)	(54.98)	(58.18)	(55.74)	
Household size	2.40	2.40	2.38	2.42	$p = 0.86$
	(1.52)	(1.60)	(1.50)	(1.46)	
Married	0.34	0.33	0.36	0.34	$p = 0.35$
Male	0.48	0.47	0.51	0.47	$p = 0.15$
Education					
Highschool degree or higher	0.95	0.94	0.95	0.95	$p = 0.74$
College degree or higher	0.41	0.41	0.40	0.40	$p = 0.84$
Post-graduate education	0.17	0.18	0.16	0.16	$p = 0.49$
Ethnicity					
Asian	0.03	0.03	0.04	0.03	$p = 0.88$
Caucasian	0.77	0.76	0.77	0.78	$p = 0.57$
Hispanic	0.03	0.04	0.03	0.03	$p = 0.47$
African American	0.07	0.08	0.07	0.08	$p = 0.83$
Other	0.02	0.03	0.02	0.02	$p = 0.39$
Tax rate in city of residence	7.32	7.36	7.31	7.30	$p = 0.36$
	(1.15)	(1.16)	(1.13)	(1.15)	
N	2998	1102	982	914	

Notes: This table presents the means and standard deviations of demographic variables in each of the three arms in our final sample. To test whether each characteristic is equally distributed across arms, we regress that characteristic on dummies for arms of the study, using OLS with robust standard errors, and report the F-test p-value for equality of across arms. Omnibus tests also show that here are no significant differences in demographics between Arm 1 vs. Arm 2 (F-test $p = 0.49$), Arm 2 vs. Arm 3 (F-test $p = 0.94$) or Arm 1 vs. Arm 3 (F-test $p = 0.36$).

Table 2: Estimates of Average θ (Weight Placed on Tax) by Experimental Arm

	(1) All	(2) $p_2 \geq 1$	(3) $p_2 \geq 5$
Std. tax avg. θ	0.261**	0.250***	0.226**
	(0.111)	(0.095)	(0.094)
Triple tax avg. θ	0.481***	0.475***	0.535***
	(0.045)	(0.039)	(0.041)
Observations	59960	58478	32810
Difference p-val	0.03	0.01	< 0.001

Notes: This table displays GMM estimates of average θ by experimental arm, applying the methodology discussed in Section 4.3. θ is defined as the "weight" that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. Column (1) uses all data, Column (2) conditions on module 2 price (p_2) being greater than 1, Column (3) conditions on module 2 price (p_2) being greater than 5. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Average θ (Weight Placed on Tax) for Different Product Valuations

	(1) Standard	(2) Triple	(3) Pooled	(1) Standard	(2) Triple	(3) Pooled
Middle p_2 bin	–0.097	0.117**	0.123**	–0.077	0.097**	0.104***
	(0.147)	(0.054)	(0.054)	(0.101)	(0.038)	(0.038)
High p_2 bin	0.115	0.147**	0.154**	0.168	0.069	0.072
	(0.185)	(0.074)	(0.074)	(0.152)	(0.053)	(0.053)
Std. Tax cons.	0.266*		0.156			
	(0.140)		(0.098)			
Triple tax cons.		0.402***	0.395***			
		(0.054)	(0.054)			
Individual fixed effects	No	No	No	Yes	Yes	Yes
Observations	40651	39378	58478	40651	39378	58478

Notes: This table displays GMM estimates of the relationship between average θ and the valuation of the good considered, applying the methodology discussed in Section 4.3. θ is defined as the "weight" that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. Columns (1)-(3) estimate the model $\bar{\theta}_{ik} = \alpha_0^{1x}\mathbf{1}_{1x} + \alpha_0^{3x}\mathbf{1}_{3x} + \alpha_{p_2\in[5,10)}\mathbf{1}_{p_2\in[5,10)} + \alpha_{p_2\geq 10}\mathbf{1}_{p_2\geq 10}$. We assume that $\alpha_{p_2\in[5,10)}$ and $\alpha_{p_2\geq 10}$ do not change across the standard and triple tax arms, but we allow for different baseline values α_0^{1x} and α_0^{3x}. Columns (4)-(6) control for individual fixed effects, estimating the model $\bar{\theta}_{ik} = \theta_i + \alpha_{p_2\in[5,10)}\mathbf{1}_{p_2\in[5,10)} + \alpha_{p_2\geq 10}\mathbf{1}_{p_2\geq 10}$. We model the two moment conditions for each arm separately, and we use the two-step GMM estimator to approximate the efficient weighting matrix for the over-identified model. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Distribution of Self-Classifying Survey Responses

	Standard	Triple
"Yes"	0.06	0.11
"Maybe a little"	0.56	0.56
"No"	0.38	0.32
Ranksum $z = 3.80$, $p < 0.001$		

Notes: Respondents were asked whether they would buy products at higher tag prices if there was no tax in the first module. The multiple-choice options were "Yes" ($R = H$), "Maybe a little" ($R = M$), or "No" ($R = L$). We report the distribution separately for the standard and triple tax arms, and test for a difference in distributions in the lower panel of the table.

Table 5: Average θ (Weight Placed on Tax) Conditional on Self-Classifying Survey Response

	(1) Standard	(2) Triple
"Yes" average θ	1.103***	0.936***
	(0.277)	(0.103)
"A little" average θ	0.436***	0.622***
	(0.110)	(0.048)
"No" average θ	–0.172	0.047
	(0.139)	(0.056)
Observations	40651	39378

Notes: This table displays GMM estimates of average θ by consumers' responses to the self-classifying survey questions, applying the methodology discussed in Section 4.3. θ is defined as the "weight" that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. Column (1) provides estimates for the standard tax arm and Column (2) provides estimates for the triple tax arm. All specifications condition on module 2 price (p_2) being greater than 1. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Lower Bound Estimates for the Expected Conditional Variance of (Weight Placed on Tax)

	Standard	Triple
Lower bound estimate	0.132	0.094
Bias (mean)	0.009	0.001
Standard Error	0.051	0.019
95% Conf. Int.	(0.054,0.251)	(0.063, 0.135)
Bias-corrected Conf. Int.	(0.049, 0.237)	(0.064, 0.136)

Notes: This table presents lower bounds for $E_{p_1,\tau}[Var[\theta|p_1\tau]]$, estimated for both the standard and triple tax arms using the methodology of Section 5. θ is defined as the "weight" that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. We compute standard errors and mean bias (Efron 1982) using the percentile (non-accelerated) bootstrap (with 1000 iterations), blocking by consumers. We compute approximate 95% confidence intervals using the unadjusted bootstrap, as well as the median bias correcting bootstrap (Efron 1987).

References

Abaluck, Jason and Jonathan Gruber, "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 2011, *101* (4), 1180–1210.

Abeler, Johannes and Simon Jäger, "Complex Tax Incentives," *American Economic Journal: Economic Policy*, 2015, *7* (3), 1–28.

Allcott, Hunt, "Consumers' Perceptions and Misperceptions of Energy Costs," *American Economic Review*, 2011, *101* (3), 98–104.

_ , "Paternalism and Energy Efficiency: An Overview," *NBER Working Paper No. 20363*, 2015.

_ **and Dmitry Taubinsky**, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 2015, *105* (8), 2501–2538.

_ **, Sendhil Mullainathan, and Dmitry Taubinsky**, "Energy Policy with Externalities and Internalities," *Journal of Public Economics*, 2014, *112*, 72–88.

Ambuehl, Sandro, B. Douglas Bernheim, and Annamaria Lusardi, "The effect of financial education on the quality of decision making," *working paper*, 2016.

Auerbach, Alan J., "The Theory of Excess Burden and Optimal Taxation," in Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Elsevier Science Publishers B. V., 1985, pp. 61–128.

Behaghel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon, "Sample Attrition Bias in Randomized Experiments: A Tale of Two Surveys," *Paris School of Economics Working Paper No. 2009-15*, 2009.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot, "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference," *American Economic Review*, September 2014, *104* (9), 2698–2735.

Benzarti, Youssef and Dorian Carloni, "What Goes Up May Not Come Down: Asymmetric Passthrough of Consumption Taxes," *working paper*, 2016.

Bernheim, B Douglas and Antonio Rangel, "Beyond Revealed Preference: Choice-theoretic Foundations for Behavioral Welfare Economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

Besley, Timothy J. and Harvey S. Rosen, "Sales taxes and prices: an empirical analysis," *National Tax Journal*, 1999, *52*, 157–178.

Bhargava, Saurabh and Day Manoli, "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

Brown, Jeffrey R., Arie Kapteyn, Erzo F.P. Luttmer, and Olivia S. Mitchell, "Cognitive Constrains on Valuing Annuities," *Journal of the European Economic Association*, forthcoming.

Bushong, Benjamin, Joshua Schwartzstein, and Matthew Rabin, "A Model of Relative Thinking," *Working Paper*, 2015.

Caplin, Andrew and Mark Dean, "The Behavioral Impications of Rational Inattention with Shannon Entropy," *NBER Working Paper No. 19318*, 2013.

_ **and** _ , "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

_ **and** _ , "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *NBER Working Paper No. 19876*, 2015.

Chetty, Raj, "The Simple Economics of Salience and Taxation," Technical Report, National Bureau of Economic Research 2009.

_ , "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018.

_ , "Behavioral Economics and Public Policy: A Pragmatic Perspective," *American Economic Review Papers and Proceedings*, 2015, *105* (5), 1–33.

_ **, Adam Looney, and Kory Kroft**, "Salience and Taxation: Theory and Evidence," *NBER working paper No. 13330*, 2007.

_ **,** _ **, and** _ , "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

_ **and Emmanuel Saez**, "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," *American Economic Journal: Applied Economics*, 2013, *5* (1), 1–31.

_ **, John N Friedman, and Emmanuel Saez**, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 2013, *103* (7), 2683–2721.

de Bartolome, Charles A. M., "Which Tax Rate do People Use: Average or Marginal?," *Journal of Public Economics*, 1995, *56* (1), 79–96.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–372.

Doyle, Joseph J. and Krislert Samphantharak, "2.00 Dollar Gas! Studying the effects of a gas tax moratorium," *Journal of Public Economics*, 2008, *92* (3-4), 869–884.

Farhi, Emmanuel and Xavier Gabaix, "Optimal Taxation with Behavioral Agents," Working Paper 21524, National Bureau of Economic Research September 2015.

Feldman, Naomi E. and Bradley J. Ruffle, "The Impact of Including, Adding, and Subtracting a Tax on Demand," *American Economic Journal: Economic Policy*, February 2015, *7* (1), 95–118.

_ **, Peter Katuscak, and Laura Kawano**, "Taxpayer Confusion: Evidence from the Child Tax Credit," *American Economic Review*, March 2016, *106* (3), 807–35.

Feldman, Naomi, Jacob Goldin, and Tatiana Homonoff, "Raisin the Stakes: Experimental Evidence on the Endogeneity of Taxpayer Mistakes," *working paper*, 2015.

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

Gabaix, Xavier, "A Sparsity Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129*, 1661–1710.

_ **and David Laibson**, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2).

Goldin, Jacob, "Optimal Tax Salience," *Journal of Public Economics*, 2015, *131*, 115 – 123.

_ **and Tatiana Homonoff**, "Smoke gets in your eyes: cigarette tax salience and regressivity," *American Economic Journal: Economic Policy*, 2013, *5* (1), 302–336.

Harberger, Arnold C., "The Measurement of Waste," *American Economic Review*, 1964, *54* (3), 58–76.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka, "Inferior Products and Profitable Deception," *Working Paper*, 2014.

Hoopes, Jeffrey, Daniel H. Reck, and Joel Slemrod, "Taxpayer Search for Information: Implications for Rational Attention," *American Economic Journal: Economic Policy*, 2015, *7* (3), 177–208.

Hossain, Tanjim and John Morgan, "...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay," *Advances in Economic Analysis and Policy*, 2006, *5* (2).

Jones, Damon and Aprajit Mahajan, "Time-Inconsistency and Saving: Experimental Evidence from Low-Income Tax Filers," *NBER Working Paper No. 21272*, 2015.

Liebman, Jeffrey B. and Richard Zeckhauser, "Schmeduling," *Working Paper*, 2004.

Lockwood, Benjamin B. and Dmitry Taubinsky, "Regressive Sin Taxes," *Working Paper*, 2015.

Lusardi, Annamaria and Olivia S. Mitchell, "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth," *Journal of Monetary Economics*, 2007, *51* (1), 205–224.

_ **and** _ , "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education," *Business Economics*, 2007, *42* (1), 35–44.

_ **and** _ , "Planning and Financial Literacy: How do Women Fare?," *American Economic Review*, 2008, *98* (2), 413–417.

_ **and** _ , "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 2014, *52* (1), 5–44.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J Congdon, "A Reduced-Form Approach to Behavioral Public Finance," *Annual Review of Economics*, 2012, *4*, 1–30.

Rees-Jones, Alex and Dmitry Taubinsky, "Heuristic Perceptions of the Income Tax: Evidence and Implications for Debiasing," *Working Paper*, 2016.

Stango, Victor and Jonathan Zinman, "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees," *Review of Financial Studies*, 2014, *27* (4), 990–1030.

Veiga, Andre and Glen Weyl, "Product Design in Selection Markets," *Quarterly Journal of Economics*, Forthcoming.

Woodford, Michael, "Inattentive Valuation and Reference-Dependent Choice," *Working Paper*, 2012.

Zimmermann, Florian and Benjamin Enke, "Correlation Neglect in Belief Formation," *Working Paper*, 2015.

Appendix (not for publication)

A Appendix to Section 2: Further Results

A.1 Identification from Aggregate Demand Data

Definition 1. Let $B_\epsilon(p,t) := [p-\epsilon, p+\epsilon] \times [t-\epsilon, t+\epsilon]$. We say that local knowledge of $D(p,t)$ is sufficient to identify the efficiency cost of a small tax change at a pre-existing tax t if for each sequence of $\{\Delta_i\}_{i=1}^{\infty}$ converging to zero there is a sequence of $\{\epsilon_i\}_{i=1}^{\infty}$ converging to zero with the property that knowledge of $B_{\epsilon_i}(p,t)$ is sufficient to identify $EB(t+\Delta_i) - EB(t)$.

Proposition 7. *Consider $\Delta EB(\Delta t|t) := EB(t+\Delta t) - EB(t)$. Suppose that $F(\theta|p,t)$ is degenerate and suppose for simplicity that utility is quasilinear.*

1. *(CLK and Chetty 2009) Suppose that either i) $F(\theta|p,t)$ does not depend on t or that ii) $t=0$. Then local knowledge of $D(p,t)$ is sufficient to identify ΔEB for a small Δt.*

2. *Suppose that $F(\theta|p,t)$ depends on t, and that $t>0$. Then local knowledge of $D(p,t)$ is not sufficient to identify excess burden or $F(\theta|p,t)$. However, full knowledge of $D(p,t)$ is sufficient to identify ΔEB for a small Δt.*

Proposition 7 shows that when $F(\theta|p,t)$ is degenerate, the demand curve $D(p,t)$ identifies welfare. In fact, when attention does not vary with the tax, the proposition shows that local knowledge of the demand curve is sufficient—a replication of CLK and Chetty et al. (2009) for the case of binary demand. The reason is that when θ is exogenous, it is given by $\frac{D_t}{D_p}$: the extent to which consumers underreact to a change in the tax relative to a change in the posted price.

When θ can depend on t, the ratio $\frac{D_t}{D_p}$ no longer identifies θ. The reason is that a change in the tax also change attention: the ratio $\frac{D_t}{D_p}$ now gives $\theta(t) + \theta'(t)$. To calculate welfare, however, it is necessary to know both $\theta(t)$ and $\theta'(t)$, as shown in Proposition 5. However, full knowledge of $D(p,t)$ is still sufficient to calculate $\theta(t)$. Intuitively, to calculate $\theta(t)$, we simply need to find the value Δp such that $D(p+\Delta p, 0) = D(p,t)$. Then $\theta(t) = \Delta p/t$, and $\theta'(t)$ can then be backed out from the demand response.

Proposition 8. *Suppose for simplicity that utility is quasilinear. Consider$\Delta EB(\Delta t|t) := EB(t+\Delta t) - EB(t)$, and let ΔEB_0 be the value of ΔEB that would be inferred from $D(p,t)$ under the assumption that $F(\theta|p,t)$ is degenerate. Then there exist $\underline{d} \leq \Delta EB_0 < \bar{d}$ such that $D(p,t)$ can be consistent with any value of $\Delta EB(\Delta t|t)$ in $[\underline{d}, \bar{d}]$. When $t=0$, and when $\bar{\theta}$ is an upper bound on the possible realizations of θ,*

$$\underline{d} = \frac{1}{2}\left(\frac{D_t(p,\Delta t)}{D_p(p,\Delta t)}\right)^2 (\Delta t)^2 D_p \tag{13}$$

$$\bar{d} = \frac{1}{2}\left(\frac{D_t(p,\Delta t)}{D_p(p,\Delta t)}\right) \bar{\theta} (\Delta t)^2 D_p \tag{14}$$

Proposition 8 shows that when there is heterogeneity in θ, knowledge of the demand curve $D(p,t)$ is not sufficient to calculate the welfare implications of taxation. To see the intuition for this result, consider the case in which $t=0$. In this case, welfare is proportional to $E[\theta|p,t]^2 + Var[\theta|p,t]$. The mean $E[\theta|p,t]$ is identified by D_t/D_p. However, $Var[\theta|p,t]$ cannot be identified at all from the demand curve D, as aggregate demands

do not provide information on the dispersion of the bias, only the extent to which it mutes the response to taxation *on average*. The variance is smallest when consumers are homogeneous, which corresponds to $\underline{d}$ in (13), and it is largest when all consumers either have $\theta = \bar{\theta}$ or $\theta = 0$, which corresponds to $\bar{d}$ in (14).

A.2 Income Effects

To generate results for income effects in the presence of bias heterogeneity, we will temporarily focus on a continuous demand model, as modeling income effects in a discrete choice model is somewhat awkward and is not a typical even in neoclassical results on efficiency costs of taxation. We assume that producer prices as fixed, as as characterizing efficiency costs with both income effects and endogenous producer prices is intractable even in the standard model (CLK).

We suppose that consumers have a utility function $U(x, y) = v(x) + v(y)$, and that they react to a tax as if it was θt. We suppose that U is the same for all consumers for simplicity, but more general results can be obtained for arbitrary distributions of U (analogous to our binary choice model with arbitrary distributions of tastes). We study a budget-adjustment rule where consumers react to the price-inclusive price of x as if it was $p + \theta t$, but overall do not misperceive the size of their budget as a consequence of the tax because they purchase x frequently and in relatively small amounts, and observe their new budget after every purchase of x. This gives rise to choices of x and y characterized by the following conditions

$$\frac{v'(x^*_\theta)}{u'(y^*_\theta)} = p + \theta t \tag{15}$$

$$y^* + (p+t)x^*_\theta = z \tag{16}$$

where z is the budget.

Now some routine algebra shows that $\theta = \frac{\frac{d(x^*_\theta)^c}{dt}}{\frac{d(x^*_\theta)^c}{dp}}$ for small t, where the compensated demand responses are defined by $\frac{d(x^*_\theta)^c}{dp} := \frac{d}{dp}x^*_\theta + \frac{d}{dz}x^*_\theta$ and $\frac{d(x^*_\theta)^c}{dt} = \frac{d}{dt}x^*_\theta + \frac{d}{dz}x^*_\theta$. It now follows by Proposition 2 of CLK that the efficiency cost corresponding to a type θ consumer is given by

$$\frac{1}{2}t^2\theta^2\frac{\varepsilon^c_{D,p}}{p+t}, \tag{17}$$

where $\varepsilon^c_{D,p}$ is the compensated price-elasticity of demand. Aggregating (17) across consumers yields the following analog to Proposition 2:

Proposition 9. *If terms of order t^3 are higher are negligible, and if consumers choose x and y according to (15) and (16), then $EB \approx \frac{1}{2}t^2\left[E[\theta|p,t]\frac{\varepsilon^c_{D,t}}{p+t} + Var[\theta|p,t]\frac{\varepsilon^c_{D,p}}{p+t}\right]$, where $\varepsilon^c_{D,t}$ is the compensated tax elasticity and $\varepsilon^c_{D,p}$ is the compensated price elasticity.*

Note that this proposition is essentially identical to Proposition 2 in the text, since the compensated elasticity equals the uncompensated elasticity with quasilinear utility. More generally, all the results in the body of the paper could be re-derived analogously. An important insight of this result is that even with income effects, efficiency costs are still zero when all consumers neglect the tax completely.

Note, however, that this result depends on the budget adjustment rule. If, instead, consumers first purchased large quantities of x, and only then spent the remainder of their income on y, then efficiency costs

would be positive even with complete neglect of taxes. The reason is that consumers would overspend on $x$in a way that would generate inefficiently low levels of consumption of y.

Concretely, the choice of x would now be characterized by

$$v'(x_\theta^*) = (p + \theta t)u'(z - (p + \theta t)) \tag{18}$$

rather than by $v'(x_\theta^*) = (p+\theta t)u'(z-(p+t))$. In this case, the choice of y would would be overly sensitive to change dt of the not-fully-salient commodity tax, in the sense that it would decrease by more than it would with respect to a salient lump sum tax of size dtx^*. This captures the intuition that inattention to taxes may cause consumers to misperceive their effective budget, and thus experience an unpleasant surprise later in time after seeing how much money they have left to purchase y.

Formally, routine algebra now shows that $\theta = \frac{\frac{dx_\theta^*}{dt}}{\frac{dx_\theta^*}{dp}}$ for small t. That is, θ is now the ratio of uncompensated demand responses rather than compensated demand responses. Now define $\theta^c := \frac{\theta\frac{dx_\theta^*}{dp} + x_\theta^*\frac{dx_\theta^*}{dz}}{\frac{d(x_\theta^*)^c}{dp}}$. In this case, Proposition 2 of CLK can now be used to show that when terms of order t^3 and higher are negligible,

$$EB \approx \frac{1}{2}t^2\left[E[\theta^c|p,t]\frac{\varepsilon_{D,t}^c}{p+t} + Var[\theta^c|p,t]\frac{\varepsilon_{D,p}^c}{p+t}\right]. \tag{19}$$

When $\theta = 0$ for all consumers, $\theta^c < 0$ and $\varepsilon_{D,t}^c < 0$, and thus (19) shows that excess burden can still be positive, formalizing the intuition that we sketched in the previous paragraph.

Note, however, that if x^*u''/u' is negligible, as is the case when expenditures arising from x are small relative to the total budget, then $\frac{d(x_\theta^*)^c}{dt} \approx \frac{dx_\theta^*}{dt}$ and $\frac{d(x_\theta^*)^c}{dp} \approx \frac{dx_\theta^*}{dp}$. Thus also $\theta^c \approx \theta$, and so efficiency costs can still be well approximated by the formula in Proposition 2 for small t. This captures the intuition that income effects are negligible for small-ticket purchases, and thus the derivations in our paper still hold for such decisions. Modifications to our main formulas are needed only when 1) the purchases are large and 2) the budget adjustment rule does not correspond to the one in (15) and (16).

A.3 Welfare with Redistributive Motives

We now consider a policymaker who aims not only to minimize efficiency costs, but also wishes to equalize wealth. We model this setting as simply as possible in this paper, but we refer the interested reader to Lockwood and Taubinsky (2015) and Farhi and Gabaix (2015) for richer models of tax salience with redistributive concerns. Lockwood and Taubinsky (2015), for example, consider a policymaker who has access to both a non-linear income tax and a (non)-salient commodity tax that he can apply to a sin good such as cigarette consumption. Analogous to the results in this section, Lockwood and Taubinsky (2015) also show that the welfare consequences of the less salient commodity tax depend on how attention to the tax covaries with income.[31]

We consider an economy in which consumers start with different levels of wealth $Z_1, \dots Z_N$, indexed by ω. We let F denote the joint distribution of (v, θ, ω), and we let $D^\omega(p,t)$ denote the demand curve of consumers with endowment Z_ω. We assume for simplicity that $D^\omega(p,0)$ and $D_p^\omega(p,0)$ do not depend on i.

[31] We remind the reader that while the Atkinson-Stiglitz theorem shows that commodity taxation should not be used with neoclassical consumers in the presence of nonlinear income taxation, this theorem does not hold when the income tax and the commodity tax are not equally salient, or when there are other biases that cause consumers to over- or under-consume the good in question (Lockwood and Taubinsky 2015).

We let F denote the joint distribution of θ, v, ω and we let H denote the marginal distribution of i. We continue assuming that consumers choose x if $v \geq p + \theta t$.

The government maximizes $W = \int g_\omega(Z_\omega + (v - p - t)\mathbf{1}_x)dF + \lambda D$, where λ is the marginal value of public funds (used for production of a public good, for example), and g_ω is the weight on the utility of consumers with wealth Z_w. Redistributive preferences are captured by g_ω decreasing in Z_ω. Similar results can be obtained by endowing consumers with utility functions $U(Z_\omega + (v - p - t)\mathbf{1}_x)$ instead of assuming exogenous given weights g_ω.

Proposition 10. *Set $\bar{g} := \int g_\omega$. For a small tax t,*

$$\begin{aligned} W(t) - W(0) \quad \approx \quad & \overbrace{\frac{t^2}{2}\left(\bar{g}\left(E[\theta|p,t]^2 + Var[\theta|p,t]\right)D_p(p,t) + Cov[g_\omega, (\theta-1)^2|p,t]D_p(p,t) - \bar{g}D_t(p,t)\right)}^{\textit{Welfare implications of misoptimization}} \\ + \quad & \underbrace{t(\lambda - \bar{g})D(p,t) + \frac{1}{2}t^2\lambda D_t(p,t)}_{\textit{Impact on public funds net of mechanical income effect}} \end{aligned}$$

Proposition 10 shows that just as excess burden is increasing in $E[\theta^2]$ and $Var[\theta]$, welfare is similarly decreasing in these two terms. Because the welfare formula reduces to the formula in Proposition 2 when $g_\omega = \lambda = 1$ for all ω, Proposition 10 is a generalization of our baseline result to the case in which the equalities $g_\omega = \lambda = 1$ do not hold.

The new insight that the more general welfare framework generates is that welfare is also increasing in the covariance between g_ω and the size of the mistake in computing bias. Because $(\theta - 1)^2$ attains its minimum at $\theta = 1$, welfare is decreasing in the extent to which the deviation from full rationality, either due to over- or underreaction to taxes, is concentrated on the low income earners.

In short, conditional on $E[\theta|p,t]$ and $Var[\theta|p,t]$, and knowledge of D_p, inferred welfare is lower when the mistake is concentrated on the poor. If consumers are over-spending on x because they are underreacting to the tax, the policymaker prefers that this over-spending is concentrated on consumers with low marginal social benefit from income.

B A More General Framework for Optimal Taxes

B.1 Welfare and Optimal Tax Formulas

We now suppose that while a consumer's perceived value from the good x is v, the actual social value from the consumer getting the good is $v - \gamma$. The wedge γ represents either externalities or internalities. For example, γ could correspond to consumers misperceiving the price of the good. We make several simplifying assumptions. First, we assume that we can partition consumers into θ types $j = 1, \ldots J$ such that type j consumers reacts to a tax t as if it was $\theta(t)t$. Second, we assume that terms of order $t^3 D_{pp}$ are negligible. Third, we assume that γ, v, θ are mutually independent.

The policymaker's objective function is to maximize

$$W(t) = \int [y - (p + t)\mathbf{1}_x + (v - \gamma)\mathbf{1}_x] + \lambda t D$$

We now characterize optimal taxes in this more general model.

Proposition 11. *Normalize $p = 1$, and define $\bar{\gamma} := E[\gamma]$, $a(t) := E[\theta|t]$ and $b(t) := E[\theta^2|t] = E[\theta|t]^2 + Var[\theta|t]$. Then*

1. $W'(t) = (\lambda - 1)D + [(\lambda - 1)t - \bar{\gamma}]D_t + \frac{b(t)+b'(t)t}{a(t)+a'(t)t}tD_p$

2. *The optimal tax t is implicitly defined by*

$$t = -\frac{(\lambda-1)(a(t)+a'(t)t)D - \bar{\gamma}(a(t)+a'(t))D_t}{(\lambda-1)(a(t)+a'(t)t)D_t + (b(t)+b'(t)t)D_t} \frac{(\lambda-1)(a(t)+a'(t)t) + \bar{\gamma}(a(t)+a'(t)t)\varepsilon_{D,t}}{(\lambda-1)(a(t)+a'(t)t)\varepsilon_{D,t} + (b(t)+b'(t)t)\varepsilon_{D,t} - (\lambda-1)(a(t)+a'(t))}$$

The general formula in part 1 of Proposition 11, which is an analogue of the kinds of general results derived in Farhi and Gabaix (2015) for continuous demand, is a more general manifestation of the forces discussed in our excess burden analysis in Section 2. Keeping in mind that $a(t) := E[\theta|t]$ and $b(t) := E[\theta^2|t] = E[\theta|t]^2 + Var[\theta|t]$, the formula shows that there are four key statistics: the mean, the variance, and how both of those change with respect to the tax. The frictions $\bar{\gamma}$ enter into the formula additively. The higher is $\bar{\gamma}$, the higher is the optimal tax t, and thus the larger the impact that the variance component of $b(t)$ has on welfare.

Part 2 of the proposition partially solves for the optimal tax to present formulas generalizing the usual "inverse elasticity" result from Ramsey taxation. To obtain intuition for the main result, we first focus on a simple case in which $\bar{\gamma} = 0$ and optimal taxes are not large. In this case, the optimal tax formula trades off the deadweight loss computed in Proposition 5 with the revenue gain (net of the mechanical effect on consumers' incomes).

Corollary 3. *When λ is close to 1 and $\bar{\gamma} = 0$,*

$$\frac{t}{1+t} = \frac{(\lambda-1)E[\theta|t]}{(E[\theta|t]^2 + Var[\theta|t])\,\varepsilon_{D,t}}$$

$$t \approx \frac{(\lambda-1)E[\theta|t]}{(E[\theta|t]^2 + Var[\theta|t])\,\varepsilon_{D,t}}$$

Just as Proposition 2 shows that the deadweight loss is increasing in both the mean and the variance of θ, Corollary 3 shows that the size of the optimal tax is decreasing in both the mean and the variance of θ.

In the presence of other (small) frictions, the tax must be adjusted to offset the other internalities and/or externalities captured by $\bar{\gamma}$. The extent to which the tax is adjusted depends on both average θ and on the variance. The lower is the average θ, the more the tax needs to be adjusted, as reflected by the $E[\theta|t]$ term in the numerator and the $E[\theta|t]^2$ in the denominator. On the other hand, the higher is the variance in θ, the greater the misallocation form increasing the tax, and thus the lower is the optimal tax.

Corollary 4. *When λ is close to 1 and $F(\theta|t)$ does not depend on t,*

$$t \approx \frac{(\lambda-1)E[\theta|t]}{(E[\theta|t]^2 + Var[\theta|t])\,\varepsilon_{D,t}} + \bar{\gamma}\frac{E[\theta|t]}{(E[\theta|t]^2 + Var[\theta|t])}$$

As a last special case for obtaining intuition, we focus on the case in which $Var[\theta|t] = 0$ for all t.

Corollary 5. *Suppose that* $Var[\theta|t] = 0$ *. Then*

$$t = \frac{\lambda - 1 + \bar{\gamma}\varepsilon_{D,t}}{(\lambda - 1)(\varepsilon_{D,t} - 1) + E[\theta|t]\varepsilon_{D,t}} \tag{20}$$

and when $\bar{\gamma} = 0$*,*

$$\frac{t}{1+t} = \frac{\lambda - 1}{(\lambda - (1 - E[\theta|t]))\varepsilon_{D,t}} \tag{21}$$

In this last special case, equation (21) provides a simple analog to the standard inverse elasticity rule of Ramsey taxation, showing that the rule is simply modified by the bias term $(1 - E[\theta|t])$.

B.2 Implications for Ramsey Taxation

The formulas derived so far are immediately transferable to the canonical Ramsey taxation models. In particular, let y be untaxed leisure, and let $x_1, \ldots x_K$ be the possible products consumers can purchase, and that, for simplicity, utility is separable in the consumption of these goods. Suppose that the government sets taxes $t_1, \ldots t_K$ on the k goods to meet a revenue target R. In this case, the value of public funds λ is determined endogenously. Set $\tau_i = t_i/p_i$ to be the tax rate.

In the standard Ramsey model, the taxes are determined by the inverse elasticity rule

$$\frac{\tau_i/(1+\tau_i)}{\tau_j/(1+\tau_j)} = \frac{\varepsilon_{D^j,t_j}}{\varepsilon_{D^i,t_i}}$$

What are the implications of tax salience? For simple intuition, suppose first that $Var[\theta|t] = 0$ and that $\bar{\gamma} = 0$. Suppose, moreover, that θ depends only on the size of the tax, so that with uniform taxes t_k, it would be identical for across the K goods. In this case, equation (21) implies that

$$\frac{\tau_i/(1+\tau_i)}{\tau_j/(1+\tau_j)} = \frac{\varepsilon_{D^j,t_j}}{\varepsilon_{D^i,t_i}} \cdot \frac{\lambda - (1 - E[\theta|p_j, \tau_j])}{\lambda - (1 - E[\theta|p_j, \tau_j])}.$$

A key implication here is that if $E[\theta|p, \tau]$ does not depend on p or τ, then the standard inverse elasticity rule continues to hold, and thus *with a fixed revenue requirement* R, taxes are identical to what they are in the standard model. Matters are different, however, if θ is endogenous to the tax. In particular, if $E[\theta|p, \tau]$ is increasing in p and/or τ, then the inverse elasticity rule becomes dampened toward uniform taxation, as consumers will be more attentive to higher taxes, and thus higher taxes generate relatively higher efficiency costs. Additionally, if $E[\theta|p, \tau]$ is increasing in p (because taxes are higher on more expensive items keeping the tax rate constant), then tax rates should be lower on more expensive products. More generally, the inverse elasticity rule is modified by how both the mean and the variance change with respect to the tax.

B.3 Salient vs. Not-Fully Salient Sales Taxes

We now consider how inattention to taxes impacts the highest attainable welfare. Building on Farhi and Gabaix (2015), we compare welfare under the not-fully salient tax t to welfare under a fully salient tax s.

Proposition 12. *1. Suppose that* $\lambda > 1$*, that* $\bar{\gamma} \geq 0$*, and that* θ *is homogeneous. Then the highest possible welfare attainable with* t *is strictly higher than the highest welfare attainable with* s*.*

2. *Suppose that $\bar{\gamma} > 0$ and that θ is heterogeneous. For λ sufficiently close to 1, the highest possible welfare attainable with s is strictly higher than the highest possible welfare attainable with t.*

The intuition is as follows. When $\bar{\gamma} = 0$ so that the purpose of taxes is to only raise revenue, less salient taxes are better because they can raise revenue in a less distortionary way. On the other hand, when $\bar{\gamma} = 1$ and $\lambda = 1$, fully salient taxes can achieve the first best, while not-fully-salient taxes cannot because different consumers will react to taxes differently.

B.3.1 An Example

Suppose that a fraction ρ have $\theta = 1$ and the rest have $\theta = 0$. Suppose also that θ is independent of v. Finally, suppose that the distribution of v is uniform in the range of taxes considered, so that the demand curve is linear in the range of taxes considered. Letting m denote the slope of the demand curve with respect to p, we now have that

$$t^* = -\frac{(\lambda-1)D - \bar{\gamma}\rho m}{(\lambda-1)\rho m + \rho m}$$

Now $W'(t) = (\lambda-1)D + \rho(\lambda-1)tm - \bar{\gamma}\rho m + \rho tm = (\lambda-1)D + \rho\lambda mt - \bar{\gamma}m$. Thus

$$\begin{aligned} W(t^*) - W(0) &= \int [(\lambda-1)(D_0 + \rho mt) + \rho\lambda mt - \bar{\gamma}\rho m] \\ &= \int [(\lambda-1)D_0 + (2\lambda-1)\rho mt - \bar{\gamma}m] \\ &= (\lambda-1)D_0 t^* + \frac{2\lambda-1}{2}\rho m(t^*)^2 - \bar{\gamma}\rho m t^* \end{aligned}$$

By the envelope theorem, $\frac{d}{d\rho}W(t^*) = \frac{2\lambda-1}{2}m(t^*)^2 - \bar{\gamma}mt^*$, and is thus positive if and only if

$$\begin{aligned} \bar{\gamma} &> (\lambda - 1/2)t^* \\ &= (\lambda - 1/2)\frac{-(\lambda-1)D + \bar{\gamma}\rho m}{(\lambda-1)\rho m + \rho m} \end{aligned}$$

From this it follows that $\frac{d}{d\rho}W(t^*) > 0$ if $\bar{\gamma}[(\lambda-1)\rho m + \rho m] > (\lambda - 1/2)(1-\lambda)D + \bar{\gamma}(\lambda - 1/2)\rho m$ or

$$\bar{\gamma} > \frac{(2\lambda-1)(1-\lambda)}{\rho m}. \tag{22}$$

Equation (22) provides conditions under which welfare is increasing in ρ, the fraction of consumers with $\theta = 1$. When $\bar{\gamma}$ is sufficiently large relative to $\lambda - 1$, it is better if more consumers are paying attention to the tax, as that reduces the inefficiencies created from some consumers over-purchasing the x and others under-purchasing it.

C Proofs of Propositions

Proof of Proposition 1 Let p_0 be the initial price and let $p(t)$ be the final price set by producers. Let x_1^* be the equilibrium quantity after the tax change, and let x_0^* be the equilibrium quantity before the tax change. The formula for excess burden is given by

$$\begin{aligned} EB(t,F) &= \overbrace{\left[\int_{v\geq p_0}(v-p_0)dF - \int_{v\geq p(t)+\theta t}(v-p(t)-t)dF\right]}^{\text{Equivalent variation in wealth for consumers}} - \overbrace{\int_{v\geq p(t)+\theta t} tdF}^{\text{Change in government revenue}} \\ &+ \overbrace{(p_0x_0^* - C(x_0^*)) - (p(t)x_1^* - C(x_1^*))}^{\text{Change in producer profits}} \\ &= \int_{v\geq p_0}(v-p_0)dF - \int_{v\geq p(t)+\theta t}(v-p(t))dF + (p_0x_0^* - C(x_0^*)) - (p(t)x_1^* - C(x_1^*)) \end{aligned} \tag{23}$$

Now by the multidimensional Leibniz rule,

$$\begin{aligned} \frac{d}{dt}\int_{v\geq p(t)+\theta t}(v-p(t))dF &= -\int\left[(\theta t)\frac{d}{dt}(\theta t + p(t))\right]dF(\theta, v|v = p(t)+\theta t) \\ &+ \int_{v\geq p(t)+\theta t}\left(-\frac{dp(t)}{dt}\right)dF(\theta,v) \\ &= -\int(\theta t)\left(\theta + \frac{dp(t)}{dt}\right)dF(\theta, v|v=p(t)+\theta t) \\ &+ \int_{v\geq p(t)+\theta t}\left(-\frac{dp(t)}{dt}\right)dF \\ &= -\ tE_F\left[\theta^2 + \theta\frac{dp(t)}{dt}|v = p(t)+\theta t\right]\int dF(\theta, v|v=p(t)+\theta t) - x_1^*\frac{dp(t)}{dt} \\ &= -tE_F[\theta]E\left[\theta + \frac{dp(t)}{dt}|v=p(t)+\theta t\right]\int dF(\theta, v|v=p(t)+\theta t) \\ &- \ tVar_F[\theta|v=p(t)+\theta t]\int dF(\theta, v|v=p(t)+\theta t) - x_1^*\frac{dp(t)}{dt} \\ &= tE_F[\theta|p(t),t]\frac{d}{dt}x_1^* + tVar_F[\theta|p(t),t]D_p(p(t),t) - x_1^*\frac{dp(t)}{dt} \end{aligned} \tag{24}$$

To arrive to the final equation in (24) from the preceding equation, we use the fact that

$$\begin{aligned} \frac{d}{dt}x_1^* &= \frac{d}{dt}\int_{v\geq p(t)+\theta t}dF \\ &= -\int\frac{d}{dt}(p(t)+\theta t)dF(\theta, v|v=p(t)+\theta t) \\ &= -\int\left(\theta + \frac{d}{dt}p(t)\right)dF(\theta, v|v=p(t)+\theta t) \end{aligned} \tag{25}$$

Next, the Envelope Theorem implies that

$$\frac{d}{dt}\left(p(t)x_1^* - C(x_1^*)\right) = x_1\frac{dp(t)}{dt}$$

Putting this together, we thus have that

$$\frac{d}{dt}EB(t,F) = -tE_F[\theta|v = p(t)+\theta t]\frac{d}{dt}x_1^* - tVar_F[\theta|v = p(t)+\theta t]D_p(p(t),t) \tag{26}$$

Proof of Proposition 2 Assuming that $E[\theta|p_1,t]$, $Var[\theta|p_1,t]$, D, and x_1^* are smooth, it follows from (26) that when F_t does not depend on t

$$\frac{d^2}{dt^2}EB(t,F) = -E_F[\theta|v = p_1+\theta t]\frac{d}{dt}x_1^* - Var_F[\theta|v = p_1+\theta t]D_p(p_1,t) + O(t) \tag{27}$$

where $O(t)$ represents all terms of order t or higher (as $t \to 0$). A Taylor expansion thus implies that when F_t does not depend on t,

$$\begin{aligned} EB(t,F) &= EB(0,F) + \Delta t\frac{d}{dt}EB(t,F)|_{t=0} + \frac{(\Delta t)^2}{2}\frac{d^2}{dt^2}EB(t,F)|_{t=0} + O((\Delta t)^3) \\ &= -\frac{1}{2}t^2\left[E_F[\theta|v = p_1+\theta t]\frac{d}{dt}x_1^* + Var_F[\theta|v = p_1+\theta t]D_p(p_1,t)\right] + O(t^3) \end{aligned} \tag{28}$$

where $O(t^3)$ represents all terms of order t^3 or higher (as $t \to 0$). Finally, because $EB(t,F_t) = EB(t,F_t) - EB(0,F_0) = EB(t,F_t) - EB(0,F_t)$, the assumption that F_t does not depend on F is without loss of generality when computing $EB(t,F_t)$. Thus formula (28) holds of how $F_{t'}$ compares to F_t for $t' \leq t$.

Proof of Proposition 3. Differentiating the expression in (26) yields

$$\begin{aligned} \frac{d^2}{dt^2}EB(t,F) &= -E_F[\theta|v = p(t)+\theta t]\frac{d}{dt}x^* - Var_F[\theta|v = p_1+\theta t]D_p(p_1,t) \\ &- t\frac{d}{dt}E_F[\theta|v = p(t)+\theta t]_{t=t_1}\frac{d}{dt}x^* - t\frac{d}{dt}Var_F[\theta|v = p_1+\theta t]|_{t=t_1}D_p \\ &- tE_F[\theta|v = p_1+\theta t]\frac{d^2}{dt^2}x^* - tVar_F[\theta|v = p_1+\theta t]\frac{d}{dt}D_p \end{aligned}$$

We now proceed under the assumption that in addition to $t(\Delta t)^2 D_{pp}$ being negligible, $t(\Delta t)^2 D_{pt}$ is negligible as well. We will show that this condition holds under the assumptions of Proposition 3 that $v \perp \theta$ and $t(\Delta t)^2 D_{pp}$ is negligible. Now $\frac{d}{dt}D_p = D_{pp}p'(t) + D_{pt}$, which implies that the term $t(\Delta t)^2\frac{d}{dt}D_p$ is negligible.

Next, note that equation (25) implies that $\frac{d}{dt}x^* = E[\theta|p,t]D_p + p'(t)D_p$. To derive $p'(t)$, note that it satisfies $D(p(t),t) = S(p(t))$, and thus $D_p p'(t) + D_t = S_p p'(t)$, which implies that $p'(t) = \frac{D_t}{S_p - D_p} = \frac{E[\theta|p,t]D_p}{S_p - D_p}$. Thus $\frac{d}{dt}x^* = E[\theta|p,t]\left(\frac{D_p}{S_p - D_p}D_p + D_p\right)$. Using similar reasoning as in the paragraph above, we can show that the term $t(\Delta t)^2\frac{d}{dt}\left(\frac{D_p}{S_p - D_p}D_p + D_p\right)$ is negligible, and thus that

$$t(\Delta t)^2 E_F[\theta|v = p_1+\theta t]\frac{d^2}{dt^2}x^* \approx t(\Delta t)^2\frac{d}{dt}E[\theta|v = p+\theta t]\frac{d}{dt}x^*.$$

Now a second order Taylor expansion shows that

$$
\begin{aligned}
EB(t+\Delta t,F)-EB(t,F) &= \Delta t\frac{d}{dt}EB(t,F)|_{t=t_1}+\frac{(\Delta t)^2}{2}\frac{d^2}{dt^2}EB(t,F)|_{t=t_1}+O((\Delta t)^3)\\
&\approx -t_1(\Delta t)E_F[\theta|v=p(t_1)+\theta t_1]\frac{d}{dt}x^*-t_1(\Delta t)Var_F[\theta|v=p(t_1)+\theta t_1]D_p(p(t_1),t_1)\\
&\quad -\frac{(\Delta t)^2}{2}\left(E_F[\theta|v=p_1+\theta t_1]\frac{d}{dt}x^*+Var_F[\theta|v=p_1+\theta t]D_p(p_1,t_1)\right)\\
&\quad -t_1(\Delta t)^2\frac{d}{dt}E[\theta|v=p+\theta t_1]\frac{d}{dt}x^*-\frac{t_1(\Delta t)^2}{2}\frac{d}{dt}Var_F[\theta|v=p_1+\theta t_1]D_p\\
&\approx -t_1(\Delta t)E_F[\theta|v=p(t_2)+\theta t_2]\frac{d}{dt}x^*\\
&\quad -t_1\Delta t\left(\frac{Var_F[\theta|v=p(t_1)+\theta t_1]+Var_F[\theta|v=p(t_2)+\theta t_2]}{2}\right)\\
&\quad -\frac{(\Delta t)^2}{2}\left(E_F[\theta|v=p(t_1)+\theta t_1]+Var_F[\theta|E_F[\theta|v=p(t_1)+\theta t_1]\right)
\end{aligned}
$$

Note that the reason the terms proportional to $\frac{d}{dt}E[\theta|p,t]$ and $\frac{d}{dt}Var[\theta|p,t]$ may not be negligible is because if θ and v are significantly correlated, then distribution of θ of the marginal consumers may depend on p and t in a signifiant way.

We now derive the simpler expression in the body of the paper under the assumption that $v \perp \theta$. Let $H(\theta)$ be the marginal distribution of θ and let $G(v)$ be the marginal distribution of v with density function. First, we show $t_1(\Delta t)^2 D_{pt}$ is negligible under this assumption, combined with the assumption that $t(\Delta t)^2 D_{pp} \approx 0$. We have

$$
\begin{aligned}
t_1(\Delta t)^2\frac{d}{dt}D_p &= t(\Delta t)^2\frac{d}{dt}\int D_p(p(t)+\theta t,0)dH(\theta)\\
&= \int (t_1(\Delta t)^2)(\theta D_{pp}+p'(t)D_{pp})dH(\theta)\\
&\approx 0
\end{aligned}
$$

Similarly, we can show that $t_1(\Delta t)^2\frac{d}{dt}\int \theta D_p dH(\theta) \approx 0$. Thus, since $E[\theta|p,t]=\frac{\int \theta D_p dH}{\int D_p dH}$, it follows that $t_1(\Delta t)^2\frac{d}{dt}E[\theta|p(t),t]|_{t=t_1}\approx 0$. Similarly, $t_1(\Delta t)^2\frac{d}{dt}E[\theta^2|p(t),t]|_{t=t_1}\approx 0$. From this it then also follows that

$$
t_1(\Delta t)^2\frac{d}{dt}Var[\theta|p(t),t]_{t=t_1} = t(\Delta t)^2\frac{d}{dt}\left(E[\theta^2|p(t),t]+E[\theta|p(t),t]^2\right)_{t=t_1}\approx 0
$$

Finally, since $\frac{d}{dt}x^*=E[\theta|p,t]\left(\frac{D_p}{S_p-D_p}D_p+D_p\right)$, it follows from above results that $t_1(\Delta t)^2\frac{d^2}{dt^2}x^*\approx 0$. Putting this all together, we thus have

$$
\begin{aligned}
\frac{d^2}{dt^2}EB(t,F) &= -E_F[\theta|v=p(t)+\theta t]\frac{d}{dt}x^*-Var_F[\theta|v=p_1+\theta t]D_p(p_1,t)\\
&\quad + \text{ terms that are negligible when mulitplied by } t(\Delta t)^2
\end{aligned}
$$

The result in the proposition now follows from the second order Taylor expansion $EB(t+\Delta t,F)-EB(t,F)=$

$\Delta t \frac{d}{dt} EB(t,F) + \frac{(\Delta t)^2}{2} \frac{d^2}{dt^2} EB(t,F) + O((\Delta t)^3)$.

Proof of Proposition 4 We write the equilibrium price as a function of n here, suppressing the dependency on t.

Under condition A,

$$\begin{aligned}
\frac{d}{dn}\int_{v\geq p(n)+\theta t}(v-p(n))dF_n &= \frac{d}{dn}\int_{v\geq p(n)+(\theta+h(n))t}(v-p(n))dF_0 \\
&= \int\left[(\theta t)\frac{d}{dn}((\theta+h(n))t+p(n))\right]dF_0(\theta,v|v=p(n)+(\theta+h(n))t) \\
&+ \int_{v\geq p(n)+\theta t}(-p'(n))\,dF(\theta,v) \\
&= \int(\theta t)(h'(n)t+p'(n))dF_0(\theta,v|v=p(n)+(\theta+h(n))t)-p'(n)D \\
&= t^2E_{F_n}[\theta|p(n),t]\left(\frac{d}{dn}E_{F_n}[\theta|p,t]|_{p=p(n)}+p'(n)\right)D_p-p'(n)D \\
&= \frac{t^2}{2}\frac{d}{dn}E_{F_n}[\theta|p,t]^2|_{p=p(n)}D_p+tE_{F_n}[\theta|p(n),t]p'(n)D_p-p'(n)D
\end{aligned}$$

where $\frac{d}{dn}E_{F_n}[\theta|p,t]|_{p=p(n)}$ denotes the derivative of $E_{F_n}[\theta|p,t]$ with respect to $n$when the price p is held constant at the value $p(n)$, and likewise for the variance operator. Since $E[\theta^2|p,t]=E[\theta|p,t]^2+Var[\theta|p,t]$, and condition A implies that $\frac{d}{dn}Var[\theta|p,t]=0$, we can replace $\frac{d}{dn}E_{F_n}[\theta|p,t]^2|_{p=p(n)}$ with $\frac{d}{dn}E_{F_n}[\theta^2|p,t]|_{p=p(n)}$ in the formula above.

Under condition B, we can rewrite the distributions as follows: Let i index "bias types" in the population, distributed according to $\tilde{H}$, with $\theta_i(n)$ the underreaction parameter for each bias type, as a function of n. By our assumptions, $\theta_i(n)$ is differentiable in n. Letting G denote the distribution of v with density function g, we now have

$$\begin{aligned}
\frac{d}{dn}\int_{v\geq p(n)+\theta t}(v-p(n))dF_n &= \frac{d}{dn}\int_{v\geq p(n)+\theta_i(n)t}(v-p(n))dG(v)d\tilde{H}(i) \\
&= \int\left[(\theta_i(n)t)\frac{d}{dn}(\theta_i(n))t+p(n))\right]g(p+\theta_i(t)t)d\tilde{H} \\
&+ \int_{v\geq p(n)+\theta t}(-p'(n))\,dF(\theta,v) \\
&= t^2\int(\theta_i(n)\theta_i'(n)+\theta_i(n)p(n))g(p)d\tilde{H}-p'(n)D+O(t^3) \\
&\approx \frac{t^2}{2}\frac{d}{dn}E_{F_n}[\theta^2|p,t]|_{p=p(n)}D_p+tE_{F_n}[\theta|p(n),t]p'(n)D_p-p'(n)D
\end{aligned}$$

where we use the fact that $t^2g(p+\theta t)=t^2g(p)+O\left(t^{k+1}\frac{\partial^k}{\partial p^k}D\right)$ for $k\geq 2$.

Finally, for producer profits, the Envelope Theorem implies that

$$\frac{d}{dn}\left(p(n)x^*-C(x^*)\right)=Dp'(n).$$

Putting this together using (23) shows that

$$\frac{d}{dn}EB(t,F_n) = -\frac{t^2}{2}\frac{d}{dn}E_{F_n}[\theta^2|p,t]|_{p=p(n)}D_p - tE_{F_n}[\theta|p(n),t]p'(n)D_p,$$

from which the first statement in the proposition follows since $p'(n) = 0$ when the elasticity of supply is infinite.

To prove the second part of the proposition, we now compute the second derivative of excess burden. We first establish several useful facts:

1. $t(\Delta n)^2 \frac{d}{dn}D_p$ is negligible
2. $t(\Delta n)^2 p''(n)$ is negligible

Under condition A, $\frac{d}{dn}D_p = D_{pp} + p'(n)h'(n)D_{pp}$. Under condition B,

$$\begin{aligned}\frac{d}{dn}D_p &= \frac{d}{dn}\int D_p(p(n)+\theta_i t,0)d\tilde{H} \\ &= \int \theta_i'(n)tD_{pp} + p'(n)D_{pp})dH(\theta)\end{aligned}$$

In both cases it follows that $t(\Delta n)^2 \frac{d}{dn}D_p$ is negligible if $t(\Delta n)^2 D_{pp}$ is negligible. Implicit differentiation also shows that $p'(n) = \frac{\frac{\partial}{\partial n}D}{S_p - D_p}$. Since $\frac{\partial}{\partial n}D = \frac{\partial}{\partial n}E[\theta|p,t]_{p=p(n)}D_p$, similar reasoning also shows that $t(\Delta t)^2 p''(n)$ is negligible. Thus

$$(\Delta n)^2\frac{d^2}{dn^2}EB \approx -t\frac{d^2}{dn^2}E_{F_n}[\theta^2|p,t]|_{p=p(n)}D_p - t\frac{d}{dn}E_{F_n}[\theta|p,t]|_{p=p(n)}p'(n)D_p.$$

Now a second-order expansion of EB around n yields

$$\begin{aligned}EB(t,F_{n+\Delta n}) - EB(t,F_n) &= (\Delta n)\frac{d}{dn}EB(t,F_n) + \frac{(\Delta n)^2}{2}\frac{d^2}{dn^2}EB(t,F_n) + O((\Delta n)^3) \\ &- \frac{t^2}{2}\left((\Delta n)\frac{d}{dn}E_{F_n}[\theta^2|p,t]|_{p=p(n)} + \frac{(\Delta n)^2}{2}\frac{d^2}{dn^2}E_{F_n}[\theta^2|p,t]|_{p=p(n)}\right)D_p \\ &\quad -t\frac{(\Delta n)^2}{2}t\frac{d}{dn}E_{F_n}[\theta|p,t]|_{p=p(n)}p'(n)D_p + O((\Delta n)^3)\end{aligned}$$

When producer prices are fixed, $p'(n) = 0$ by definition, and

$$(\Delta n)\frac{d}{dn}E_{F_n}[\theta^2|p,t] + \frac{(\Delta n)^2}{2}\frac{d^2}{dn^2}E_{F_n}[\theta^2|p,t] = E_{F_{n+\Delta n}}[\theta^2|p,t] - E_{F_n}[\theta^2|p,t] + O(\Delta n)^3,$$

from which the statement in the proposition follows.

Proof of Proposition 5 Combining Propositions 4 and 5, we have

$$\frac{d}{dt}EB(t,F) \approx -E[\theta^2|p,t]tD_p - \frac{t^2}{2}\frac{d}{dt}E[\theta^2|p,t]D_p.$$

Taking the second derivative, and ignoring the terms proportional to $\frac{d}{dt}D_p$ since those are negligible (when multiplied by $t(\Delta t)^2$) by the reasoning in Propositions 4 and 5, we have

$$
\begin{aligned}
(\Delta t)^2 \frac{d^2}{dt^2} EB(t,F) &\approx (\Delta t)^2 \left(-E[\theta^2|p,t]D_p - t\frac{d}{dt}E[\theta^2|p,t]D_p \right) \\
&- (\Delta t)^2 t \frac{d}{dt}E[\theta^2|p,t]D_p - (\Delta t)^2 \frac{t^2}{2}\frac{d^2}{dt^2}E[\theta^2|p,t]D_p
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
(\Delta t)\frac{d}{dt}E[\theta^2|p,t]|_{t=t_1} + \frac{(\Delta t)^2}{2}\frac{d^2}{dt^2}E[\theta^2|p,t]|_{t=t_1} &= E[\theta^2|p,t_2] - E[\theta^2|p,t_1] + O((\Delta t)^3) \\
(\Delta t)E[\theta|p,t_1] + (\Delta t)^2\frac{d}{dt}E[\theta|p,t]_{t=t_1} &= (\Delta t)E[\theta|p,t_2] + O((\Delta t)^3) \\
(\Delta t)^2 E[\theta^2|p,t_1] &= (\Delta t)^2 E[\theta^2|p,t_2] + O((\Delta t)^3)
\end{aligned}
$$

Putting this together, we have

$$
\begin{aligned}
EB(t+\Delta t,F) - EB(t,F) &= \Delta t\frac{d}{dt}EB(t,F)|_{t=t_1} + \frac{(\Delta t)^2}{2}\frac{d^2}{dt^2}EB(t,F)|_{t=t_1} + O((\Delta t)^3) \\
&\approx -\left(t_1(\Delta t) + \frac{(\Delta t)^2}{2} \right) \left(E[\theta|p,t_2]^2 + Var[\theta|p,t_2] \right) D_p \\
&- \frac{1}{2}t^2 \left(E[\theta^2|p,t_2] - E[\theta^2|p,t_1] \right)|_{t=t_1} D_p
\end{aligned}
$$

Proof of Corollary 2 We have $\frac{d}{dt}D = \left(E[\theta|p,t] + \frac{d}{dt}E[\theta|p,t] \right) D_p$, and thus

$$
\begin{aligned}
\frac{d}{dt}EB(t,F) &\approx -E[\theta^2|p,t]tD_p - \frac{t^2}{2}\frac{d}{dt}E[\theta^2|p,t]D_p \\
&= -E[\theta|p,t]^2 tD_p - Var[\theta|p,t]tD_p - \frac{t^2}{2}\frac{d}{dt}\left(E[\theta|p,t]^2 + Var[\theta|p,t] \right) D_p \\
&= -E[\theta|p,t]t\left(E[\theta|p,t] + \frac{d}{dt}E[\theta|p,t] \right) D_p - t\left(Var[\theta|p,t] + \frac{t}{2}\frac{d}{dt}Var[\theta|p,t] \right) D_p \\
&= -E[\theta|p,t]tD_t - t\left(Var[\theta|p,t] + \frac{t}{2}\frac{d}{dt}Var[\theta|p,t] \right) D_p
\end{aligned}
$$

Next,

$$\begin{aligned}
\frac{d^2}{dt^2}EB(t,F) &\approx -E[\theta|p,t]D_t - \left(Var[\theta|p,t] + t\frac{d}{dt}Var[\theta|p,t]\right)D_p \\
&\quad -t\frac{d}{dt}E[\theta|p,t]D_t - E[\theta|p,t]t\frac{d}{dt}\left(E[\theta|p,t] + \frac{d}{dt}E[\theta|p,t]\right)D_p \\
&\quad -t\left(\frac{d}{dt}Var[\theta|p,t] + \frac{t}{2}\frac{d^2}{dt^2}Var[\theta|p,t]\right)D_p + O(tD_{pp}) \\
&= -E[\theta|p,t]D_t - \left(Var[\theta|p,t] + t\frac{d}{dt}Var[\theta|p,t]\right)D_p \\
&\quad -t\frac{d}{dt}E[\theta|p,t]D_t - \frac{t}{2}\frac{d}{dt}\left(E[\theta|p,t]^2 + Var[\theta|p,t]\right)D_p \\
&\quad -\frac{t}{2}\frac{d}{dt}Var[\theta|p,t]D_p - tE[\theta|p,t]\frac{d^2}{dt^2}\left(E[\theta|p,t] + Var[\theta|p,t]\right)D_p + O(tD_{pp})
\end{aligned}$$

Now since the approximations $E[\theta|p,t_2] - E[\theta|p,t_1] \approx \Delta t\frac{d}{dt}E[\theta|p,t]|_{t=t_1}$ and $Var[\theta|p,t_2] - Var[\theta|p,t_1] \approx \Delta t\frac{d}{dt}Var[\theta|p,t]|_{t=t_1}$ are valid, the terms are $(\Delta t)^2\frac{d^2}{dt^2}E[\theta|p,t]$ and $(\Delta t)^2\frac{d^2}{dt^2}Var[\theta|p,t]$ are negligible. Next, note that

$$(\Delta t)E[\theta|p,t_1] + \frac{1}{2}(\Delta t)^2\frac{d}{dt}E[\theta|p,t]_{t=t_1} = (\Delta t)\frac{E[\theta|p,t_1] + E[\theta|p,t_2]}{2} + O((\Delta t)^3) \quad (29)$$

$$(\Delta t)Var[\theta|p,t_1] + \frac{1}{2}(\Delta t)^2\frac{d}{dt}Var[\theta|p,t]_{t=t_1} = (\Delta t)\frac{Var[\theta|p,t_1] + Var[\theta|p,t_2]}{2} + O((\Delta t)^3) \quad (30)$$

$$(\Delta t)D_p\left(E[\theta^2|p,t_2] - E[\theta^2|p,t_1]\right) = (\Delta t)^2\frac{d}{dt}E[\theta^2|p,t]|_{t=t_1} + O((\Delta t)^3) \quad (31)$$

$$(\Delta t)D_p\left(Var[\theta|p,t_2] - Var[\theta|p,t_1]\right) = (\Delta t)^2\frac{d}{dt}Var[\theta|p,t]|_{t=t_1} + O((\Delta t)^3) \quad (32)$$

Now,

$$\begin{aligned}
EB(t+\Delta t,F) - EB(t,F) &= \Delta t\frac{d}{dt}EB(t,F)|_{t=t_1} + \frac{(\Delta t)^2}{2}\frac{d^2}{dt^2}EB(t,F)|_{t=t_1} + O((\Delta t)^3) \\
&\approx -t_1(\Delta t)\left(\left(E[\theta|p,t_1] + \frac{d}{dt}E[\theta|p,t]|_{t=t_1}\right)D_t + \left(Var[\theta|p,t_1] + \frac{d}{dt}Var[\theta|p,t]|_{t=t_1}\right)D_p\right) \\
&\quad - \frac{(\Delta t)^2}{2}\left(E[\theta|p,t_1]D_t + Var[\theta|p,t_1]D_p\right) \\
&\quad - \frac{1}{2}(t(\Delta t) + t)\left(\frac{d}{dt}Var[\theta|p,t]|_{t=t_1}\right)D_p \\
&\quad - \frac{t}{2}(\Delta t)^2\frac{d}{dt}E[\theta^2|p,t]|_{t=t_1}D_{p,}
\end{aligned}$$

from which the the statement in the proposition follows by making the substitutions in (29)-(32).

Proof of Proposition 6 *Step 1.* We first show that $Var[\theta|p,\tau] \geq Var\left[\phi|p,\tau\right]$, where $\phi = \frac{\log(1+\theta\tau)}{\tau}$. Define a new function $\mathcal{J}(\theta) = \theta - \phi(\theta)$. Note that $\mathcal{J}' = 1 - 1/(1+\theta\tau)$; that is, $\mathcal{J}$ is increasing. Now consider

$$\mathcal{F}(k) = Var[\theta - k\mathcal{J}(\theta)|p,t] = E\left[\left(\theta - k\mathcal{J}(\theta) - E[\theta - k\mathcal{J}(\theta)|p,t]\right)^2|p,t\right]$$

The derivative with respect to k is

$$\begin{aligned}\frac{d}{dk}\mathcal{F}(k) &= -2E\left[\left(\left(\theta - k\mathcal{J}(\theta) - E[\theta - k\mathcal{J}(\theta)|p,t]\right)\left(\mathcal{J}(\theta) - E[\mathcal{J}(\theta)|p,t]\right)\right] \\ &= -2Cov\left[\left(\theta - k\mathcal{J}(\theta) - E[\theta - k\mathcal{J}(\theta)|p,t], \mathcal{J}(\theta) - E[\mathcal{J}(\theta)|p,t]|p,t\right] \\ &< 0\end{aligned}$$

where the inequality follows because the two random variables in the covariance operator are both increasing in θ and thus must have positive covariance. This shows that $\mathcal{F}(k)$ is decreasing in k. But note that by definition, $\mathcal{F}(0) = Var[\theta|p,\tau]$ and $\mathcal{F}(1) = Var[\phi|p,\tau]$ since $\theta - \mathcal{J}(\theta) = \phi(\theta)$.

Intuitively, we are taking a distribution of θ, and we are modifying it by subtracting from each outcome θ a function $J(\theta)$ that is increasing in θ. Thus we are modifying the distribution in a way that pulls in the highest realizations the most toward zero—exactly the kind of operation that reduces variances.

Step 2. For each consumer marginal at price p and tax τ, and with survey response $R = r$, define $\bar{\phi}(r,p,\tau) = E[\phi|r,p,\tau]$. Note that for each pair (p,τ), the distribution of ϕ is a mean preserving spread of the distribution of $\bar{\phi}$. Thus $E\left[Var\left[\phi|p,\tau\right]\right] \geq E[Var[\bar{\phi}|p,\tau]]$.

Step 3. Let $\mu(p,\tau) = E[\phi(\theta)|p,\tau]$, and let G be the induced distribution of (p,τ). Then

$$\begin{aligned}E[Var[\bar{\phi}|p,\tau]] &= \int\left[\sum_r Pr(R=r|p,\tau)(\bar{\phi}(r,p,\tau) - \mu(p,\tau))^2\right]dG(p,\tau) \\ &= \sum_r \int Pr(R=r|p,\tau)(\bar{\phi}(r,p,\tau) - \mu(p,\tau))^2 dG(p,\tau)\end{aligned}$$

Now

$$Pr(R=r)\int Pr(R=r|p,\tau)(\bar{\phi}(r,p,\tau) - \mu(r,p,\tau))^2 dG(p,\tau) \tag{33}$$

$$= E_G\left[\left(Pr(R=r|p,\tau)^{1/2}\right)^2\right] E_G\left[\left(Pr(R=r|p,\tau)^{1/2}(\bar{\phi} - \mu)\right)^2\right] \tag{34}$$

$$\geq \left(E_G\left[Pr(R=r|p,\tau)^{1/2} \cdot Pr(R=r|p,\tau)^{1/2}(\bar{\phi} - \mu)\right]\right)^2 \tag{35}$$

$$= \left[\int(\bar{\phi}(r,p,\tau) - \mu(p,\tau))Pr(R=r|p,\tau)dG\right]^2 \tag{36}$$

$$= \left[Pr(R=r)E[\bar{\phi}|R=r] - Pr(R=r)E[\mu|R=r]\right]^2 \tag{37}$$

$$= Pr(R=r)^2(E[\bar{\phi}|R=r] - E[\mu|R=r])^2. \tag{38}$$

In the computations above, line (34) follows from line (33) by definition. Line (35) follows from line (34) by the Cauchy-Schwarz inequality. And line (37) follows from line (36) because by definition,

$$\begin{aligned} E[\bar{\phi}|R=r] &= \frac{\int \bar{\phi} Pr(R=r,p,\tau)dG}{\int Pr(R=r,p,\tau)dG} \\ &= \frac{\int \bar{\phi} Pr(R=r,p,\tau)dG}{Pr(R=r)} \end{aligned}$$

and

$$\begin{aligned} E[\mu|R=r] &= \frac{\int \mu Pr(R=r,p,\tau)dG}{\int Pr(R=r,p,\tau)dG} \\ &= \frac{\int \mu Pr(R=r,p,\tau)dG}{Pr(R=r)} \end{aligned}$$

This implies

$$\int Pr(R=r|p,\tau)(\bar{\phi}-\mu)^2 dG \geq Pr(R=r)(E[\bar{\phi}|R=r]-E[\mu|R=r])^2$$

and thus

$$E[Var[\bar{\phi}|p,\tau]] \geq \sum_r Pr(R=r)(E[\phi(\theta)|R=r]-E[\mu|R=r])^2. \tag{39}$$

Proof of Proposition 7 With minor abuse of notation, we let $\theta(t)$ denote the (homogeneous) θ, as a function of t.

Part 1. Note that $\theta = D_t/D_p$. Now

$$\begin{aligned} EB'(t) &= (1-\theta)tD_t(p,t) \\ &= [1-D_t(p,t)/D_p(p,t)]tD_t(p,t) \end{aligned}$$

Thus if $D(p,t')$ is known for all values $t' \in [t, t+\Delta t]$, $EB(t+\Delta)-EB(t)$ is identified by $\int_{t'=t}^{t'=t+\Delta} EB'(t')dt'$.

Part 2. We show that $EB'(t)$ cannot be identified if $D(p,t)$ is known only in a small neighborhood around (p,t). Because $EB'(t) = (1-\theta(t))tD_t(p,t)$, it is necessary to identify $\theta(t)$. Concretely, suppose that we observe D in the neighborhood $\mathbb{R}^+ \times [t_1,t_2]$, with $t_1 > 0$. The data is rationalized if there exist functions ψ and $\theta(t)$ such that $D(p,t) = \psi(p+\theta(t)t)$ for all p and $t \in [t_1,t_2]$. Now consider one such pair of functions ψ and θ. We show that these are not uniquely determined. This $\theta(t)$ is not uniquely determined. In particular, consider $\tilde{\theta}(t) = \theta(t) - \epsilon t_1/t$, and $\tilde{\psi}(x) = \psi(x+\epsilon)$. But $\psi(p+\theta(t)t) = \tilde{\psi}(p+\tilde{\theta}(t)t)$, and thus θ is not uniquely identified by the data. In particular, note that while $\tilde{\theta}(t_1) < \theta(t_2)$, it is also true that $\tilde{\theta}'(t) > \theta'(t)$ for $t > t_1$. Intuitively, by making the slope of $\theta(t)$ steeper while making the base level lower, we are able to imitate the demand response to the tax. Again, the core principle here is that $D_t/D_p = \theta(t) + \theta'(t)$, and thus while the sum of the level and slope is identified, these are not identified separately.

On the other hand, full knowledge of D is sufficient to identify $\theta(t)$ for each t. Simply let $\Delta p(t)$ be the value for which $D(p+\Delta p, 0) = D(p,t)$. Then $\theta(t) = \Delta p/t$.

Proof of Proposition 8 First, we show that every demand curve $D(p,t)$ can be rationalized by assuming that $F(\theta|v,t)$ is degenerate. In particular, consider a function $\psi(p)$ such that $\psi(p) = D(p,0)$ for all p. Now to derive our candidate θ, start with $\nu(p,t)$ satisfying $\psi(p+\nu(p,t)t) = D(p,t)$ for all p,t. By definition, $p+\theta(v(p,t),t)t = v(p,t) = p+\nu(p,t)t$, which implicitly defines the $\theta(v,t)$ that, together with ψ, rationalizes $D(p,t)$. By definition, the valuation of a consumer marginal at (p,t) is given by $v = p+\nu(p,t)t$. Thus the data is rationalized by ψ and θ satisfying $\theta(\nu(p,t)t+p,t) = \nu(p,t)$. In this case, $EB'(t) = -(\theta t)D_t$.

Alternatively rationalize $D(p,t)$ by a distribution in which a consumer has $\theta = \bar{\theta}$ with probability $q(v,t)$, and $\theta = 0$ with probability $1-q(v,t)$. Set $\tilde{q}(p,t)$ to satisfy $\tilde{q}(p,t)D(p+\bar{\theta}t,0) + (1-\tilde{q}(p,t))D(p,0) = D(p,t)$. Note that because $D(p,0) \geq D(p,t) \geq (p+\bar{\theta}t,0)$ by definition, $\tilde{q}(p,t) \in [0,1]$. Now a consumer with $\theta = \bar{\theta}$ is marginal at (p,t) if $v = p+\bar{\theta}t$. Thus $q(v,t)$ rationalizes $D(p,t)$ if $q(p+\bar{\theta}t,t)D(p+\bar{\theta}t,0) + (1-q(p+\bar{\theta}t,t))D(p,0) = D(p,t)$. In this case $EB'(t) = -\bar{\theta}tD_t$. Now by construction, $\nu(p,t) < \bar{\theta}$, and thus $EB'(t)$ is higher in the case with heterogeneous θ.

Finally, to establish the bounds for $t = 0$ and $\Delta t \to 0$, note that $EB(\Delta t) \to -\frac{1}{2}t^2(E[\theta|p,0]^2 + Var[\theta|p,0])D_p$ as $\Delta t \to 0$. Now $E[\theta|p,0]$ is pinned down by $D_t(p,0)/D_p(p,0)$. But the variance is highest when all consumers are either $\theta = \bar{\theta}$ or $\theta = 0$.

Proof of Proposition 10

$$W(t) = \int_{v<p+\theta t} g_\omega Z_\omega d\tilde{F} + \int_{v\geq p+\theta t} g_\omega(Z_\omega - p - t + v)dF + \int_{v\geq p+\theta t} t\lambda dF$$

Analogous to the strategy for excess burden, define $\tilde{W}(t,\tilde{F})$ to be the welfare at a tax t given a distribution $\tilde{F}(\theta,v,\omega)$ that does not depend on t. Let $\tilde{F}(\theta,v,\omega) = F(\theta,v,\omega|t)$ here. Then

$$\begin{aligned}
\frac{d}{dt}\tilde{W} &= \int g_\omega \left[\theta Z_\omega - \theta(Z_\omega + \theta t - t)\right] d\tilde{F}(v,\theta,\omega|v = p+\theta t) \\
&- \int_{v\geq p+\theta t} g_\omega d\tilde{F} - t\lambda D_t(p,t) + \lambda D(p,t) \\
&= t\int g_\omega \theta(1-\theta) d\tilde{F}(v,\theta,\omega|v = p+\theta t) \\
&- \int_{v\geq p+\theta t} g_\omega d\tilde{F} - t\lambda D_t(p,t) + \lambda D(p,t) \\
&= -t\sum_\omega g_\omega E[\theta(1-\theta)|p,t,\omega]D_p^\omega(p,t) \\
&- \int_{v\geq p} g_\omega d\tilde{F} + \int_{p\leq v\leq p+\theta t} g_\omega d\tilde{F} + t\lambda D_t(p,t) + \lambda D(p,t) \\
&= -t\sum_\omega g_\omega E[\theta(1-\theta)|p,t,\omega]D_p^\omega(p,t) \\
&- \sum_\omega g_\omega D^\omega(p,0) - t\sum_\omega g_\omega E[\theta|p,t,\omega]D_p^\omega(p,t) + t\lambda D_t(p,t) + \lambda D(p,t) \\
&= -\sum_\omega g_\omega D^\omega(p,0) - t\sum_\omega g_\omega E[2\theta - \theta^2|p,t,\omega]D_p^\omega(p,t) + t\lambda D_t(p,t) + \lambda D(p,t) \\
&= -\bar{g}D(p,0) - tCov[g_\omega, 2\theta - \theta^2]D_p - 2t\bar{g}D_t^\omega + t\bar{g}E[\theta^2]D_p + t\lambda D_t + \lambda D \\
&= -\bar{g}D + t(\lambda - \bar{g})D_t + t\bar{g}E[\theta^2|p,t]D_p + \lambda D - tCov[g_\omega, 2\theta - \theta^2]D_p
\end{aligned}$$

Thus

$$\frac{d^2}{dt^2}\tilde{W} = (\lambda-\bar{g})D_t + (\lambda-\bar{g})D_t + \bar{g}E[\theta^2|p,t]D_p - Cov[g_\omega, 2\theta-\theta^2]D_p + O(t)$$

A second order Taylor expansion thus implies that

$$\begin{aligned} W(t)-W(0) &= t(\lambda-\bar{g})D(p,0) + \frac{t^2}{2}\bar{g}\left(E[\theta|p,t]^2 + Var[\theta|p,t]\right)D_p - \frac{t^2}{2}Cov[g_\omega, 2\theta-\theta^2]D_p + t^2(\lambda-\bar{g})D_t + O(t^3) \\ &= t(\lambda-\bar{g})D(p,t) + \frac{t^2}{2}\bar{g}\left(E[\theta|p,t]^2 + Var[\theta|p,t]\right)D_p - \frac{t^2}{2}Cov[g_\omega, 2\theta-\theta^2]D_p + \frac{t^2}{2}(\lambda-\bar{g})D_t + O(t^3) \end{aligned}$$

D Additional Econometric Results

D.1 Appendix to Section 4.3: Robustness to Selection on Subject Comprehension

Let $\pi \in \{0,1\}$ denote whether the person passes the quiz question or not. Let η denote the characteristics associated with passing. Continue letting X denote the vector of covariates of θ. Let $\phi = \frac{\log(1+\theta\tau)}{\tau}$

Proposition 13. *1. Assume that for both the standard and triple tax arms (*C=1x *or* C=3x*),* $E[y_{ik}|\pi = 1, C, X] = E[y_{ik}|\pi = 1, C]$*. Then for the standard and triple tax arms (*C=1x *or* C=3x*),*

$$E[y_{ik}|\pi=1,C,X] = E[y_{ik}|\pi=1,C,\theta_{ik}=0] + E\left[\phi|\pi=1,C,X\right]$$

2. Assume A1 and set $m(X) = E[y_{ik}|\pi = 1, C{=}0x, X]$*. Then for* C=1x *or* C=3x*,*

$$\frac{E[y_{ik}|\pi=1,C,X] - m(X)}{\tau_i} = E[\phi|\pi=1,C,X]$$

3. (Behaghel et al., 2009) Assume that $Pr(\pi = 1|\eta, C{=}3x) \leq Pr(\pi = 1|\eta, C{=}1x)\ \forall\ \eta$*. Then*

$$\begin{aligned} &\int E[\phi|\eta,\pi=1,C{=}3x,X]dF(\eta|\pi=1,C{=}1x,X) - \int E[\phi|\eta,\pi=1,C{=}1x,X]dF(\eta|\pi=1,C{=}1x,X) \\ \geq &\frac{Pr(\pi=1|C{=}3x)}{Pr(\pi=1|C{=}1x)}E[\phi|\pi=1,C{=}3x,X] - E[\phi|\pi=1,C{=}1x,X] \end{aligned} \tag{40}$$

Part 1 of the proposition states that when order effects do not interact with the θ covariates X, no further assumptions are necessary to study how θ *changes* with some vector of covariates X. For example, no assumptions are necessary to measure the averages *differences* in θ across the self-sorting survey questions. Of course, our estimates are for the subgroup generated by $\pi = 1$ in condition C=1x or 3x. This is analogous to the local average treatment effect in an IV regression.

Part 2 of the proposition says that under assumption A1, selection on quiz questions does not confound questions about the average value of θ conditional on covariates X and an experimental condition C. In particular, our analysis of individual differences is not confounded by differential pass rates between the no tax and tax arms. Roughly, A1 holds when 1) Characteristics η associated with passing do not interact

with order effects and 2) conditional on characteristics, different experimental conditions do not generate differences in order effects. Again, the estimates are for the subgroups generated by $\pi = 1$ in condition C=1x or 3x, rather than for the full group of study participants taking part in the experiment.

Both parts 1 and 2 of the proposition derive results for average θ conditional on an experimental condition. Part 3 of the proposition—which is derived in Jones and Mahajan (2015) and Behaghel et al. (2009)—deals with the question of how to compare average θ across conditions. Here, we use an additional monotonicity condition to derive a lower bound for the difference in average θ between conditions C=3x and C=1x. In essence, the monotonicity condition states that any subject who did not pass the comprehension check in the standard tax arm also would not pass the comprehension check in the triple tax arm.

Intuitively, the "worst case scenario" for the lower bound is when the study participants who pass in condition C=1x but not in condition C=3x have $\theta = 0$. The lower bound corresponds to this scenario, in which case $E[\theta|\pi = 1, \text{C=3x}, X]$ must be deflated by the ratio $\frac{Pr(\pi=1|\text{C=3x})}{Pr(\pi=1|\text{C=1x})}$ to derive the treatment effect of higher taxes for the types of study participants who pass in condition 1x. Again, the treatment effect here is the average treatment effect on the types of study participants who pass in condition C=1x in the experiment, rather than the average treatment effect on all types in the experiment. We implement this approach in Section 4.4.

Implementation:

To implement the lower-bound estimate (40), we estimate three moment conditions: The first two are the moment conditions (5) and (6) for study participants who pass the comprehension questions—these give us estimates of $E[\theta|\pi = 1, \text{C=3x}]$ and $E[\theta|\pi = 1, \text{C=1x}]$. The third moment condition employs the full sample to estimate $\frac{Pr(\pi=1|\text{C=3x})}{Pr(\pi=1|\text{C=1x})}$. We use these estimate to derive the lower-bound (40), and we use the delta method to obtain standard errors.

D.2 Within-Subject Estimation of Endogenous Attention

Let X^1_{ik} denote whether $p_2 \in [5, 10)$ for consumer i's kth product. Similarly, define X^2_{ik} to be an indicator for $p_2 \in [5, 10)$ for consumer i's kth product. For $\phi_{ik} = \frac{\log(1+\theta_{ik}\tau_i)}{\tau_i}$, we model

$$E[\phi_{ik} \quad |\mathbf{1}_{p_2\in[5,10)}, \alpha_{p_2\geq 10}\mathbf{1}_{p_2\geq 10}] = \quad \alpha_i + \alpha_{p_2\in[5,10)}X^1_{ik} + \alpha_{p_2\geq 10}X^2_{ik}$$

and

$$E[y_{ik}] \quad = \quad b_i + b_1 X^1_{ik} + b_2 X^2_{ik} + \phi_{ik}.$$

We set $\bar{\phi}_i = \frac{1}{20}\sum_k \phi_{ik}$, $\bar{\bar{y}}_i = \frac{1}{20}\sum_k y_{ik}$, $\overline{X^h_i} = \frac{1}{20}\sum_k X^h_{ik}$. From this, it follows that

$$\begin{aligned} E\left[\frac{y_{ik} - \bar{\bar{y}}_i}{1 - \mathbf{1}_{tax} + \tau_i \mathbf{1}_{tax}}\right] \quad &= \quad \frac{b_1 X^1_{ik} - b_1\overline{X^1_i}}{1 - \mathbf{1}_{tax} + \tau_i \mathbf{1}_{tax}} + \frac{b_2 X^2_{ik} - \overline{b_2 X^2_i}}{1 - \mathbf{1}_{tax} + \tau_i \mathbf{1}_{tax}} \\ &+ \quad \alpha_{p_2\in[5,10)}(X^1_{ik} - \overline{X^1_i})\mathbf{1}_{tax} + \alpha_{p_2\geq 10}(X^2_{ik} - \overline{X^2_i})\mathbf{1}_{tax} \end{aligned} \tag{41}$$

To estimate the parameters, we proceed as before with method of moments, replacing the theoretical moment in (41) with the empirical moment. Note that equation (41) does not contain any of the terms α_i,

and simply identifies the terms $\alpha_{p_2\in[5,10)}$ and $\alpha_{p_2\geq 10}$ using only within-consumer variation. This is analogous to estimating a linear fixed-effects model with the standard demeaning fixed-effects estimator.

D.3 Further Details for the Lower-Bound Estimation

The statistic with which we approximate the lower bound from Proposition 6 is

$$\sum_{r\in\{L,M,H\}} Pr(R=r)(\bar{\theta}_r - E[\tilde{\mu}|R=r])^2 \tag{42}$$

To estimate (42), we estimate each $\bar{\theta}_r$ using the empirical moment version of the left-hand-side of (7). We estimate $\tilde{\mu}(p_1,\tau)$ using the empirical moment counterpart of

$$E\left[\frac{y_{ik} - E[y_{ik}|\text{no tax arm}, p_1^{ik}\in\mathbf{p}(p_1),\tau_i\in\boldsymbol{\tau}(\tau)]}{\tau_i}|p_1^{ik}\in\mathbf{p}(p_1),\tau_i\in\boldsymbol{\tau}(\tau)\right] \tag{43}$$

where $E[y_{ik}|C=0\text{x}]$ denotes the average change in valuations that occurs between module 1 and module 2, and is identified from the no tax arm. We estimate $E[\tilde{\mu}|R=r]$ by computing the empirical average over all pairs (p,τ) associated with $R=r$ in the dataset. For concreteness, we construct the estimator for the standard tax arm. The estimator for the triple tax arm is analogous.

We estimate $PR(R=r)$ by $\widehat{Pr(R=r)} := \frac{1}{N}\sum \mathbf{1}_{R_i=r}$, where N is the number of participants in the standard tax arm, and $\mathbf{1}_{R_i=r}$ is an indicator that consumer i's response was r. We estimate $\bar{\theta}_r$ by

$$\widehat{\bar{\theta}}_r = \frac{1}{N_r}\sum_{i,k}\left[\frac{y_{ik} - E[y_{ik}|\text{no tax arm}]}{\tau_i}|R_i=r\right].$$

where N_r is the number of consumer-product pairs associated with $R_i=r$. We estimate $E[y_{ik}|\text{no tax arm}, \mathbf{p}\times\boldsymbol{\tau}]$, the average order effect in the no tax arm, by

$$m(\mathbf{p}\times\boldsymbol{\tau}) := \frac{1}{|\mathbf{p}\times\boldsymbol{\tau}|_{no\ tax}}\sum_{(p_1^{ik},\tau_i)\in\mathbf{p}\times\boldsymbol{\tau}} y_{ik}\mathbf{1}_{\text{no tax}}$$

where $|\mathbf{p}\times\boldsymbol{\tau}|_{no\ tax}$ is the number of observations (p_1,τ) in the interval $\mathbf{p}\times\boldsymbol{\tau}$ in the no tax arm. We estimate $\tilde{\mu}$ by

$$\widehat{\tilde{\mu}(p_1,\tau)} := \frac{1}{|\mathbf{p}(p_1)\times\boldsymbol{\tau}(\tau)|}\sum_{(p_1^{ik},\tau_i)\in\mathbf{p}\times\boldsymbol{\tau}}\frac{y_{ik} - m(\mathbf{p}\times\boldsymbol{\tau})}{\tau_i} \tag{44}$$

Clearly, $\widehat{\tilde{\mu}(p_1,\tau)}$ is an unbiased estimate of $E\left[\frac{y_{ik}-E[y_{ik}|\text{no tax arm},\mathbf{p}(p_1),\boldsymbol{\tau}(\tau)]}{\tau_i}|\mathbf{p}(p_1),\boldsymbol{\tau}(\tau)\right]$. We now end by showing that this is an unbiased estimate of $\tilde{\mu}$. To see this, note that assumption A2 implies that in the standard tax arm,

$$E\left[y_{ik}|\theta_{ik},R_j,\mathbf{p},\boldsymbol{\tau}\right] = E[y_{ik}|\text{no tax arm},\mathbf{p},\boldsymbol{\tau}] + E[\log(1+\theta_{ik}\tau_i)|\mathbf{p},\boldsymbol{\tau}]$$

from which the conclusion follows by rearrangement.

Finally, to estimate $E[\mu|R=r]$, we simply take the average of $\widehat{\mu_{ik}}$ over all observations associated with $R=r$ in the standard tax arm. We will call this $\widehat{E[\mu|R=r]}$. Our estimate of the variance bound is now

$$\sum_{r\in\{L,M,H\}} \widehat{Pr(R=r)} \left(\widehat{\bar{\theta}_r} - \widehat{E[\mu_{ik}|R=r]}\right)^2 \tag{45}$$

By construction, our estimates of $Pr(R=r)$,$\bar{\theta}_r$, and $\tilde{\mu}$ are all unbiased. Note, however, that (45) is not an unbiased estimate of the lower bound because any residual noise terms in our estimates of the moments are squared and then averaged. We estimate this mean bias with the same bootstrap procedure that we use to compute the standard errors.

E Additional Empirical Analyses and Robustness Checks

E.1 Further Tests of Module 2 Differences

Table A1: Testing for Module 2 Differences by Experimental Arm

	(1) OLS	(2) 0.25 Quantile	(3) 0.5 Quantile	(4) 0.75 Quantile
1x Arm	–0.09	–0.03	–0.03	–0.09
	(0.10)	(0.12)	(0.08)	(0.09)
3x Arm	–0.03	–0.05	–0.04	0.21
	(0.10)	(0.12)	(0.08)	(0.16)
Observations	59960	59960	59960	59960

Notes: This table tests for differences in module 2 willingness to pay for products by experimental arm. Column 1 reports estimates from an OLS regression. Columns (2)-(4) report 0.25, 0.5, and 0.75 quantile regressions. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.2 Demographic covariates

We analyze how θ covaries with standard demographics provided by ClearVoice—race, age, educational attainment, marital status—as well as the three additional covariates that we collected in our experiment, described below:

Household Income. Participants were also asked to state their household income. We analyze the data by income quartiles, the cutoffs for which are 28k, 50k, and 82k, which match almost exactly to the 2010 US census data.[32]

Ability to compute taxes / Numeracy. Immediately after the survey question about the sales tax rate, consumers were asked to compute the sales tax (in absolute terms) on an \$8 (non-tax-exempt) item. We code answers as correct if consumers provide the correct answer using their perceived sales tax rate. For example, if the true sales tax rate is 6%, but the consumer thinks that it is 7%, then an answer is coded as being correct if it is less than 1 cent from \$0.56. Consumers were asked to answer this question in the format of \$0.56. However, as with the question about sales tax beliefs, not all consumers followed the instructions. Some consumers seemed to have entered their answers in the format of \$8.56 instead of \$0.56. Other consumers seem to have entered their answers as 56 instead of \$0.56. For consumers whose answers are between 8 and 12 (about 10% of consumers), we recode answers by subtracting 8, as we think it is implausible that anyone would think that the tax on an \$8 item would be greater than \$8. For consumers whose answers are above 20,

[32] According to the 2010 census, the quartile thresholds are 25k, 50k, 90k.

we recode their answers by dividing by 100, as these consumers most likely entered their answers in number of cents rather than dollars. Our results are robust to simply excluding consumers with answers above 8. Overall, accuracy was very high, with 73% of consumers giving the right answer. That underreaction persists despite this high level of accuracy shows that consumers are either deliberately choosing not to compute the taxes, or are simply forgetting to think about taxes when determining their willingness to pay. This idea that consumers seem to make "bad" decisions despite knowing how to make "good" ones is broadly consistent with the results reported in Ambuehl et al. (2016) and Zimmermann and Enke (2015) for other types of financial decisions.

Financial Sophistication. We use the "Big Three" financial literacy questions (Lusardi and Mitchell, 2008, 2014). The three multiple choice questions test for understanding of interest rates, inflation, and risk diversification.[33] We code participants as financially sophisticated if they answer all three questions correctly. Overall, 49% of consumers in our final sample answered all three questions correctly.[34] The three questions are are as follows:

1. Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow? a) More than $102 b) Exactly $102 c) Less than $102 d) Do not know

2. Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account? a) More than today b) Exactly the same as today c) Less than today d) Do not know

3. Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund." a) True b) False c) Do not know

[33]Previous work has shown that financial literacy is associated with mistakes in other domains, including incurring overdraft fees (Stango and Zinman, 2014), incorrectly valuing annuities (Brown et al., forthcoming), and not saving enough for retirement (Lusardi and Mitchell, 2007a,b).

[34]Our measure of tax numeracy and financial sophistication are strongly correlated. Financially sophisticated consumers have a 12 percentage point grater likelihood of correctly answering the tax computation question ($p < 0.01$).

Table A2: Average θ (Weight Placed on Tax) by Demographics

	(1) Standard	(2) Triple	(3) Pooled
Compute Tax Correctly	0.123	0.132	0.138 *
	(0.204)	(0.084)	(0.084)
Financially Sophisticated	0.439 *	0.206 **	0.201 **
	(0.224)	(0.090)	(0.090)
Income Quartile 2	–0.058	0.006	–0.003
	(0.277)	(0.115)	(0.115)
Income Quartile 3	0.237	0.123	0.127
	(0.291)	(0.121)	(0.120)
Income Quartile 4	0.066	0.247 *	0.239 *
	(0.295)	(0.134)	(0.132)
Age	–0.022 ***	–0.010 ***	–0.009 ***
	(0.007)	(0.003)	(0.003)
Male	–0.026	–0.044	–0.040
	(0.210)	(0.085)	(0.085)
Married	–0.183	–0.130	–0.135
	(0.230)	(0.092)	(0.092)
College Degree	–0.159	0.046	0.037
	(0.218)	(0.095)	(0.094)
Asian	–0.607	–0.101	–0.092
	(0.525)	(0.235)	(0.237)
Caucasian	0.453	0.008	0.018
	(0.321)	(0.147)	(0.145)
Hispanic	0.737	–0.303	–0.227
	(0.517)	(0.253)	(0.252)
African American	0.664	0.008	0.001
	(0.458)	(0.191)	(0.190)
Observations	38010	36790	54643

Notes: This table displays method of moments estimates of the relationship between average θ and demographic covariates. θ is defined as the "weight" that consumers place on the sales tax, with $\theta = 0$ corresponding to complete neglect of the tax and $\theta = 1$ corresponding to full optimization. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.3 Replication of Main Results Without Excluding Study Participants Failing Comprehension Questions

Table A3: Estimates of Average θ (Weight Placed on Tax) by Experimental Arm

	(1) All	(2) $p^2 \geq 1$	(3) $p^2 \geq 5$
Std. tax avg. θ	0.064	0.107	0.146*
	(0.108)	(0.086)	(0.085)
Triple tax avg. θ	0.276***	0.292***	0.376***
	(0.040)	(0.032)	(0.033)
Observations	84460	82009	44918

Notes: This table replicates Table 2, but does not exclude study participants who failed comprehension checks. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Average θ (Weight Placed on Tax) for Different Product Valuations

	(1) Standard	(2) Triple	(3) Pooled	(1) Standard	(2) Triple	(3) Pooled
Middle p_2 bin	–0.003	0.138***	0.145***	0.023	0.116***	0.124***
	(0.133)	(0.047)	(0.047)	(0.096)	(0.035)	(0.034)
High p_2 bin	0.159	0.194***	0.201***	0.330**	0.148***	0.150***
	(0.169)	(0.063)	(0.063)	(0.138)	(0.049)	(0.049)
Std. tax cons.	0.117		0.037			
	(0.126)		(0.087)			
Triple tax cons.		0.221***	0.214***			
		(0.046)	(0.045)			
Fixed effects	No	No	No	Yes	Yes	Yes
Observations	54503	54988	82009	54503	54988	82009

Notes: This table replicates Table 3, but does not exclude study participants who failed comprehension checks. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Average θ (Weight Placed on Tax) Conditional on Self-Classifying Survey Response

	(1) Standard	(2) Triple
"Yes" average θ	0.616***	0.625***
	(0.229)	(0.081)
"A little" average θ	0.289***	0.410***
	(0.097)	(0.040)
"No" average θ	–0.246*	–0.027
	(0.126)	(0.045)
Observations	54503	54988

Notes: This table replicates Table 5, but does not exclude study participants who failed comprehension checks. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.4 Replication of Main Results with OLS Regressions

As an alternative to our GMM approach, one could instead proceed with a simple OLS framework given by:

$$E[y] = \alpha_0 + \alpha_1 \tau + \alpha_2 \tau X + \beta_1 X. \tag{46}$$

Because the linear model is misspecified when underreaction is endogenous to the tax rate, an estimate $\hat{\alpha}_1$ obtained from equation (46) is not a consistent estimate of $E[\theta_{ik}]$. In particular, the OLS estimate $\hat{\alpha}_1$ will depend on how much less consumers underreact to large taxes than to small taxes. To take a concrete illustration, we estimate an average θ of approximately 0.25 and 0.48 in the standard and triple tax arms, respectively, and thus obtain an average θ of 0.37 in the pooled sample. If we simply estimate (46) using the OLS estimator, however, we get an $\hat{\alpha}_1$ of 0.49—an estimate that is higher than either average and does not have a clear economic interpretation. While this problem can be reduced by allowing for different coefficients on the tax rate across the normal and triple tax conditions, there remains natural variation in tax rates within each experimental arm. This variation induces the same endogeneity concern—although the variation in natural tax rates is smaller, and thus the bias induced by this variation is less dramatic in magnitude.

While these concerns lead us to prefer the GMM approach as a primary specification, our qualitative results are also obtained when estimating (46) using OLS, as we show below.

Table A6: Estimates of Average θ (Weight Placed on Tax) by Experimental Arm

Dependent variable: $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$

	(1) All	(2) $p_2 \geq 1$	(3) $p_2 \geq 5$
tax × *standard*	0.254**	0.252***	0.202**
	(0.105)	(0.088)	(0.082)
tax × *triple*	0.487***	0.479***	0.532***
	(0.043)	(0.038)	(0.039)
Observations	59960	58478	32810

Notes: This table replicates Table 2, implemented through an analogous OLS procedure. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Average θ (Weight Placed on Tax) for Different Product Valuations

Dependent variable: $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$

	(1) Standard	(2) Triple	(3) Pooled	(1) Standard	(2) Triple	(3) Pooled
tax × Middle p_2 bin	–0.127	0.102**	0.109**	–0.059	0.085**	0.093***
	(0.133)	(0.051)	(0.051)	(0.097)	(0.036)	(0.036)
tax × High p_2 bin	0.055	0.135*	0.133*	0.074	0.061	0.050
	(0.164)	(0.070)	(0.071)	(0.133)	(0.048)	(0.048)
tax × std. tax arm	0.286**		0.177*			
	(0.133)		(0.094)			
tax × triple tax arm		0.415***	0.412***			
		(0.052)	(0.051)			
Fixed effects	No	No	No	Yes	Yes	Yes
Observations	40651	39378	58478	40651	39378	58478

Notes: This table replicates Table 3, implemented through an analogous OLS procedure. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Average θ (Weight Placed on Tax) Conditional on Self-Classifying Survey Response

Dependent variable: $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$

	(1) Standard	(2) Triple
tax × Yes	1.068***	0.923***
	(0.281)	(0.096)
tax × A little	0.435***	0.627***
	(0.104)	(0.047)
tax × No	–0.182	0.047
	(0.122)	(0.052)
Observations	40651	39378

Notes: This table replicates Table 5, implemented through an analogous OLS procedure. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.5 Replication of Main Results Excluding Study Participants Not Understanding the BDM Mechanism

Table A9: Estimates of Average θ (Weight Placed on Tax) by Experimental Arm

	(1) All	(2) $p^2 \geq 1$	(3) $p^2 \geq 5$
Std. tax avg. θ	0.277**	0.274***	0.262***
	(0.126)	(0.103)	(0.094)
Triple tax avg. θ	0.524***	0.513***	0.595***
	(0.052)	(0.044)	(0.046)
Observations	46540	45372	25658

Notes: This table replicates Table 2, dropping study participants who failed comprehension checks about the BDM mechanism.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Average θ (Weight Placed on Tax) for Different Product Valuations

	(1) Standard	(2) Triple	(3) Pooled	(1) Standard	(2) Triple	(3) Pooled
Middle p_2 bin	–0.153	0.156**	0.172***	–0.078	0.123***	0.132***
	(0.155)	(0.061)	(0.061)	(0.115)	(0.041)	(0.041)
High p_2 bin	0.253	0.229***	0.256***	0.183	0.095*	0.102*
	(0.206)	(0.083)	(0.082)	(0.171)	(0.054)	(0.054)
Std. tax cons.	0.296*		0.103			
	(0.159)		(0.105)			
Triple tax cons.		0.410***	0.394***			
		(0.061)	(0.060)			
Fixed effects	No	No	No	Yes	Yes	Yes
Observations	31319	30363	45372	31319	30363	45372

Notes: This table replicates Table 3, dropping study participants who failed comprehension checks about the BDM mechanism. All specifications condition on $p_2 \geq 1$. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Average θ (Weight Placed on Tax) Conditional on Self-Classifying Survey Response

	(1) Standard	(2) Triple
"Yes" average θ	0.844***	0.984***
	(0.298)	(0.113)
"A little" average θ	0.572***	0.661***
	(0.127)	(0.054)
"No" average θ	–0.219*	0.066
	(0.131)	(0.064)
Observations	31319	30363

Notes: This table replicates Table 5, dropping study participants who failed comprehension checks about the BDM mechanism. Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm. All specifications condition on $p_2 \geq 1$. Standard errors, clustered at the subject level, reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F Items Used in the Study

Product	**Amazon.com price (as of Feb 2015)**	**Amazon.com Product Description**
RainStoppers 68-Inch Oversize Windproof Golf Umbrella	$12.61	This RainStoppers 68" oversize golf umbrella is large enough to cover three or more people. Umbrella frame constructed with fiberglass shaft and ribs for maximum stability. Canopy is made of 190T Nylon fabric. Complete with a foam non slip handle. Matching sleeve included. Length when closed is 43".
Energizer AA Batteries max Alkaline 20-Pack	$11.15	energizer AA max alkaline batteries 20 pack super fresh, Expiration Date: 2024 or better. Packed in original Energizer small box 4 batteries per box x 5 boxes total 20 batteries.
Glad OdorShield Tall Kitchen Drawstring Trash Bags, Fresh Clean, 13 Gallon, 80 Count	$12.79	Glad OdorShield Tall Kitchen Drawstring Trash Bags backed by the power of Febreze are tough, reliable trash bags that neutralize strong and offensive odors for lasting freshness. These durable bags are great for use in the kitchen, home office, garage, and laundry room.
Admiral Blue 100% Cotton Bath Towel - 27 x 52 Inches	$14.99	There isn't much that's better than stepping out of a refreshing shower and wrapping yourself in the soft, Luxury Bath Towels. Now you can have that feeling every single day. It won't just be a treat anymore; it'll be your way of life. These extra-absorbent 100% cotton towels can be just hanging around waiting for you, ready to fulfill their duty in making you feel pampered. Not only practical but also stylish, these towels will also add a fashionable and luxurious touch to your bathroom.

Martex Egyptian Cotton Hand Towel with Dry-Fast (French Blue)	$6.79	Martex is one of the oldest and most trusted names in bath products. This towel is made of loops of 100% Egyptian cotton which offers the absorbency and quality of this fine extra-long-staple fiber. The towel offers DryFast Technology. Enjoy a broad color palette to compliment any bathroom decor.
Pilot G2 Retractable Premium Gel Ink Roller Ball Pens, Fine Point, Black Ink, Dozen Box (31020)	$11.89	Discover the smooth writing and comfortable G2, America's #1 Selling Gel Pen*. G2 gel ink writes 2X longer than the average of branded gel ink pens**. The G2 product line includes four point sizes, fifteen color options, and multiple barrel styles to suit every situation and personality. It is the only gel pen that offers this level of customization–because after all, pens aren't one size fits all.
Scotch-Brite Heavy Duty Scrub Sponge 426, 6-Count	$7.73	O-Cel-O™ sponges and Scotch Brite scrubbers are truly a fashion-meets-function success story. The highly absorbent and durable sponges come in different sizes and scrub levels for the various surfaces around the home. Their assorted colors and patterns follow the current fashion trends to create the perfect accent in any room.
Febreze Fabric Refresher Spring & Renewal Air Freshener, 27 Fluid Ounce	$4.94	When it comes to your home, you should never settle for less than fresh. Febreze Fabric Refresher is the first step to total freshness in every room. The fine mist eliminates odors that can linger in fabrics and air, leaving behind nothing but a light, pleasing scent. With Febreze Fabric Refresher, uplifting freshness is a simple spray away.
Microban Antimicrobial Cutting Board Lime Green - 11.5x8 inch	$8.99	The Microban cutting board from Uniware is the perfect cutting board for the health conscious. The cutting board has a soft grip with handle and is dishwasher safe. The cutting board can be reversible, use on both sides, and is non-porous, non absorbent. The rubber grips prevents slipping on countertop. Doesn't dull knives, juice-collecting groove. Microban is the most trusted antimicrobial product protection in the world. Built-In defense that inhibits the growth of stain and odor causing bacteria, mold, and mildew. Always works to keep the cutting board cleaner between cleanings. Lasts throughout the lifetime of the cutting board. Size: 11.5"x8" Color: Lime Green.
Nordic Ware Natural Aluminum Commercial Baker's Half Sheet	$11.63	Nordic Ware's line of Natural Commercial Bakeware is designed for commercial use, and exceeds expectations in the home. The durable, natural aluminum construction bakes evenly and browns uniformly, while the light color prevents overbrowning. The oversized edge also makes getting these pans in and out of the oven a cinch. Proudly made in the USA by Nordic Ware
Gain with FreshLock HE Original Liquid Detergent, 100 Fl Oz	$9.97	The scent of Gain Original liquid laundry detergent brings a lively scent to your laundry room. Powerful Lift & Lock Technology lifts away dirt and stains so you can lock in the amazing scent you love. With bursts of citrus, a green twist, and just enough floral fragrance, you'll wish laundry day came more often.

Rubbermaid Configurations Folding Laundry Hamper, 23-inch, Natural (FG4D0602NATUR)	$12.99	Rubbermaid Configurations Folding Laundry Hamper, 23-inch, Natural (FG4D0602NATUR). Makes it easy to add hamper space to any Rubbermaid Configurations Kit. Collapses for easy storage. Neutral two-tone canvas is breathable and stylish. Coordinates with other items in Rubbermaid Configurations collection. For nearly 80 years, Rubbermaid has represented innovative, high-quality products that help simplify life. Recognized as a “Brand of the Century” for its impact on the American way of life.
Scotch Precision Scissor, 8-Inches (1448)	$5.44	Scotch Precision 8" Scissors come with the finest quality stainless steel blades for a sharp edge and long cutting life. These scissors also comes with a soft grip handles for ease of use. Great for everyday cutting needs. Comes with a limited lifetime warranty.
Clorox Company 00450 Gw All Purpose Cleaner, 32-Ounce	$8.09	Cuts through grease, grime and dirt as well as traditional cleaners. Spray on counters, appliances, stainless steel, sealed granite, chrome, cook top hoods, sinks and toilets. Made from plants and minerals, 99 percent natural, so it leaves no harsh chemical fumes or residue.
Rubbermaid Easy Find Lid Medium Value Pack Food Storage Containers	$10.20	The Rubbermaid Easy Find Lids Medium Value Pack includes (2) 3.0 cup Easy Find Lid containers measures 7" x 7" x 2.3"and (1) 5.0 cup Easy Find Lid container measures 7" x 7" x 3.4". The number one unmet need for food storage is container and lid organization. With Rubbermaid’s new Easy Find Lids you’ll find storage and organization a breeze! The Easy Find Lids snap together as well as snap to the bases for easy storage. The Easy Find Lids and bases also nest together making storage in a cabinet or a drawer much more efficient. Easy Find Lids are square in shape and allow for easy of stacking when placed on shelves or in the refrigerator. Rubbermaid Easy Find Lids also feature a super clarified base which takes the guessing out of what’s inside and allows you to see what’s inside quickly and easily. Rubbermaid Easy Find Lids and bases are also microwave, freezer, and dishwasher safe.
Rubbermaid Lunch Blox medium durable bag - Black Etch	$10.47	The Rubbermaid 1813501 Lunch Blox medium durable bag - Black Etch is an insulated lunch bag designed to work with the Rubbermaid Lunch Blox food storage container system. The bag is insulated to achieve the maximum benefit of Blue Ice blocks and keep your food cold. The bag features a bottle holder, side pocket, comfort-grip handle and removable shoulder strap. The lunch Blox bag is durable and looks good for both the professional bringing their lunch to work or the kid taking their lunch to school.
Libbey 14-Ounce Classic White Wine Glass, Clear, 4-Piece	$12.99	Great for any party, this set includes four 14-ounce clear classic white wine glasses which match perfectly with the classic collection by libbey. The glasses are dishwasher safe and made in the USA.
Fulcrum 20010-301 Multi-Flex LED Task Light and Book Light	$9.47	The Multi Flex Light is an all-purpose book light, task light or travel light

Envision Home Microfiber Bath Mat with Memory Foam, 16 by 24-Inch, Espresso	$10.82	Enjoy spa luxury at home with the Envision Home Microfiber Bath Mat, featuring memory foam! Designed to absorb water like a sponge and help protect floors from damaging puddles of water, your feet will love stepping on to this soft cushion of memory foam encased in super-absorbent microfiber. The Microfiber Bath Mat starts with fibers that are split down to microscopic level, resulting in tiny threads that love to absorb every drop of water. Because of this increased surface area, this microfiber mat can collect more water than an ordinary bath mat. Plus, it dries unbelievably fast.The soft memory foam interior provides a comfortable and warm place to stand, or when kneeling to bathe a child or pet, preventing aches and pains. The seams across the mat allow for it to be easily folded for storage, or simply hang it from the convenient drying loop. It is available in three colors to compliment your personal décor and style – Cream, Celestial and Espresso.Caring for your Microfiber Bath Mat is easy; simply toss it in the washing machine with cold water and a liquid detergent and then place in the dryer on a low heat setting. The Microfiber Bath Mat is just one of the many impressive items offered in the Envision Home Collection. Designed to make it easier to take care of the home, our innovative, high-value and superior-quality products provide cleaning, kitchen, bath, laundry and pet solutions to solve life's little dilemmas.
Carnation Home Fashions Hotel Collection 8-Gauge Vinyl Shower Curtain Liner with Metal Grommets, Monaco Blue	$8.99	Protect your favorite shower curtain with our top-of-the-line Hotel Collection Vinyl Shower Curtain Liner. This standard-sized (72" x 72") liner is made with an extra heavy (8 gauge), water repellant vinyl that easily wipes clean. With metal grommets along top of the liner to prevent tearing. Here in Monaco Blue, this liner is available in a variety of fashionable colors. With its wonderful features and fashionable colors, this liner could also make a great shower curtain

Note: These are Amazon.com prices as they were displayed to, and documented by, our research assistant in February 2015. Prices may vary over time or by geographic region.