

Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment*

Dmitry Taubinsky[†] (Job Market Paper) and Alex Rees-Jones

December, 2015

Abstract

A rapidly growing literature shows that consumers can be inattentive to complex or not-fully-salient financial incentives. This literature typically estimates “the average mistake,” and uses representative agent models to analyze the economic implications. This paper shows, theoretically and empirically, that accounting for heterogeneity in mistakes is crucial in positive and normative analysis. Focusing on consumer underreaction to not-fully-salient sales taxes, we show theoretically that 1) individual differences in underreaction generate inefficiency in the resulting allocation of the taxed good, 2) the variation of underreaction across the income distribution affects the regressivity of the tax burden, and 3) the variation of underreaction across different tax rates affects the distortions to demand resulting from tax changes. To empirically assess the importance of these issues, we implement an online shopping experiment in which 3000 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) High income earners are less likely to underreact to taxes than low income earners, and thus the financial burden of misoptimization falls disproportionately on the poor. 3) Tripling existing sales tax rates roughly doubles consumers’ attention to taxes, which implies that raising taxes increases deadweight loss through an additional “debiasing channel.” 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 in other domains.

*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, Matthew Rabin, Emmanuel Saez, Andrei Shleifer, Jeremy Tobacman, Glen Weyl, Michael Woodford, Danny Yagan, and audiences at Berkeley, Carnegie Mellon SDS, CESifo Behavioral Economics Meeting, Cornell, Haas (marketing), Harvard, Stanford, and Wharton 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.

[†]Corresponding and first author. Email: dmitry.taubinsky@berkeley.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, 2014), 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, quantitative analysis demonstrating the implications of consumer mistakes for policy design typically relies on representative agent models.

Relatively little is known about how mistakes vary across consumers and across economic incentives, and how this variation impacts economic and policy analysis. We argue that understanding this variation is crucial. The magnitude of individual differences in the propensity to misoptimize 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. When there are individual differences in mistakes, whether the rich or the poor are most likely to make the mistakes matters: the more the financial burden of misoptimization falls on the low-income consumers, the more it hinders redistributive goals. And in addition to how mistakes vary between consumers, how they vary with economic incentives matters: this variation creates a new “debiasing” channel by which incentives shape consumer choice and its economic efficiency.

Focusing on the concrete setting of consumers underreacting¹ to not-fully-salient sales taxes, this paper combines a theoretical framework with a new experimental design to study the variation in consumers’ mistakes and the implications for tax policy. We present new evidence that there is significant heterogeneity in people’s underreaction to sales taxes, and that this heterogeneity has striking implications for the welfare consequences of taxation.

To formalize our arguments about the importance of variation in mistakes for policy calculations, we begin with a model—building on and generalizing Chetty (2009a) and Chetty, Looney and Kroft (2009, henceforth CLK)—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 fully incorporate its presence into their purchasing decisions. Allowing for arbitrary heterogeneity in both consumers’ valuations for the products and consumers’ underreaction to the tax, our first theoretical result is a generalization of the canonical Harberger (1964) formula. As in CLK, we find that the efficiency cost of imposing a small tax in a previously untaxed market is decreasing in the mean of consumer underreaction to sales tax. However, we additionally show that inefficiency is increasing in the variance of underreactions, to a degree of equal quantitative importance. The result arises because heterogeneity in underreactions generates misallocation of products to consumers. When

¹For expositional clarity, we use the term “underreaction” to refer to consumers’ mistakes, because both previous work as well the evidence in this paper show that this is the most common mistake. However, our formal model allows for both under- and over-reaction to sales taxes.

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 are heterogeneous in their underreaction, purchasing decisions depend on both their valuation of the good *and* on their propensity to ignore the tax. This misallocation induced by sorting on behavioral types thus creates a new inefficiency that offsets the gains from underreaction that are obtained in the absence of heterogeneity. 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 when there is misallocation due to heterogeneity in underreaction.

Next, we extend our analysis to study the efficiency costs of increasing pre-existing taxes. We show that in addition to the mean and variance of underreaction, the new key determinant of efficiency costs is the change in underreaction that occurs when taxes are increased. If increases in the tax rate “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.

Of course, policymakers may also be concerned with how the financial burden of a tax is shared between high- and low-income households. Moving beyond efficiency cost calculations, we extend our cost of taxation framework to consider a simple variation of the models in Lockwood and Taubinsky (2015) and Farhi and Gabaix (2015), in which the policymaker also has redistributive preferences. We show that with redistributive concerns, welfare is decreasing in the extent to which low-income households are most prone to tax mistakes. Intuitively, overall welfare is lower when the financial burden of mistakes falls more heavily on the lower-income households.

Taken together, these theoretical results show that empirical estimates of the variation in mistakes are crucial for welfare analysis. Our final theoretical results show, however, that 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 3000 consumers—approximately matching the US adult population on household income, gender, age, and educational attainment—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. The design affords us greater statistical power to estimate

underreaction to taxes than existing datasets, as well as new richness—both within subject and across demographic groups— for studying variation in underreaction.

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 there is no variation in underreaction. We find that in the standard tax condition study participants react to the taxes as if they are only 25% of their size. That is, a 1 dollar increase in the tax produces approximately the same change in demand as a 25 cent increase in the posted price of a product. 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 at least 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).

We next move on to study how underreaction to the tax covaries with income. We find a negative association between household income and underreaction to the sales taxes, with consumers in the fourth quartile of the income distribution approximately twice as attentive to taxes as consumers in the first quartile of the income distribution.² This difference is partially explained by differences in numeracy and financial savvy between the high and low income consumers, but persists even when controlling for (proxies for) these variables.³ At the same time, we also establish that incorrect beliefs are only a minor sources of the mistakes we measure: on average, consumers’ beliefs are not biased, and 73% of consumers know their sales tax rate within 0.5 percentage points.

Our results about observable covariates of underreaction establish, qualitatively, the existence of individual differences. In the last part of our empirical analysis, we quantify these individual differences by computing a lower bound for the variance of underreaction.⁴ 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 underreaction to the tax. We bound this statistic through a novel combination of a “self-

²See Campbell (2006), Beshears et al. (2012), Stango and Zinman (2014) for similar findings of how lower household income is predictive of mistakes in mortgage refinancing, 401(k) allocations, and incurring overdraft fees.

³Our result about numeracy is consistent with, e.g., Brown et al. (2013), who show that numeracy leads less biased annuity valuations, and Gerardi et al. (2013), who show that low numeracy leads to mortgage defaults. See Lusardi and Mitchell (2014) for a further review. Lack of financial literacy has been shown to predict mistakes in other domains, including incurring overdraft fees (Stango and Zinman, 2014), incorrectly valuing annuities (Brown et al., 2013), and not saving for retirement (Lusardi and Mitchell, 2007a,b). For previous work establishing this relationship between financial literacy and household income, see, e.g., van Rooij et al. (2012), as well as Lusardi and Mitchell (2014) for a review.

⁴We compute a bound, rather than a point estimate, because as we discuss, the noise distribution in the data unfortunately precludes precise estimates of second or higher order moments.

classifying” survey question and experimental behavior, in a way that requires no assumptions about truth-telling or metacognition. We utilize multiple-choice survey questions asking study participants whether they would be willing to buy at higher posted prices if there were no taxes in the experiment. We find that their responses—“Yes”, “Somewhat”, “No”—are highly predictive of underreaction to taxes. On average, the approximately 10% of consumers answering “Yes” do not underreact to the taxes, the approximately 55% answering “Somewhat” react to taxes as if they were only half their true size, and those answering “No” do not react to the taxes at all. Building on the fact that the true distribution of θ must be a mean-preserving spread of the distribution of conditional averages by survey response,⁵ we derive an econometrically tractable lower-bound for $E_{p,t}[Var[\theta|p, t]]$ that uses moments of θ associated with the different self-classified groups. 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 a few literatures. At a broad level, the paper contributes to a growing theoretical and empirical literature in “behavioral public economics” (see Mullainathan et al. 2012 and Chetty 2015 for reviews, and Farhi and Gabaix 2015 for a general theoretical framework). Our contribution to this literature is to emphasize the importance of studying heterogeneity in mistakes for obtaining empirically grounded, robust welfare estimates. In our own recent and concurrent work, we have touched on the identification challenges posed by individual differences in mistakes (Allcott and Taubinsky, 2015), the need to measure whether the mistakes are more or less likely to be made by the poor (Lockwood and Taubinsky, 2015), and the challenges of relying on locally-estimated elasticities when consumers are boundedly rational and attention is endogenous to incentives (Allcott, Mullainathan and Taubinsky, 2014).⁶ This paper is the first, to our knowledge, to combine a theoretical framework with an integrated experiment to explicitly bring all three facets of variation in mistakes to the forefront, and to demonstrate the significant quantitative importance of these issues. As we further discuss in section 7, our framework for analyzing variation in mistakes can serve as a template for empirical analysis for other psychological biases and domains of behavior.

More concretely, our work relates to the nascent literature on tax salience and tax misunderstanding. Building on Chetty, Looney, and Kroft (2009) and Finkelstein (2009)—who study average underreaction in a representative agent framework—we theoretically and empirically extend the literature on tax salience by showing that the variation in mistakes implies that the efficiency cost of not-fully-salient taxes may be significantly higher than previously thought. Together with our con-

⁵This general idea behind this approach is not unique to our setting. See Hendren (2013) for an application of this approach to classifying private information in the health insurance market. In conceptually similar approaches to sorting respondents into “behavioral types,” Handel and Kolstad (2015) combine survey questions with insurance choice data, and Rees-Jones (2015) combines self-classifying survey questions with forecasts of estimated utility models.

⁶See 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.

current work on income tax misperceptions (Rees-Jones and Taubinsky, 2015), our results about the association of tax mistakes with numeracy and financial literacy also contribute to a deeper understanding of the mechanisms and characteristics leading to these mistakes, and suggest a role for financial savvy as a measurable characteristic that predicts responses to tax incentives.⁷

Third, 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), as the predictions of these models rely on assumptions about heterogeneity, as well as how imperfect processing depends on the size of the shrouded attribute. 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. Our estimates can thus help guide the quantitative predictions of these models.

Fourth, our work contributes to the literature on boundedly rational value computation (see, e.g., Woodford, 2012; Gabaix, 2014; 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 clean demonstrations in a naturalistic setting of imperfect processing of a financial attribute responding to economic incentives.⁸

The paper proceeds as follows. Section 2 presents our theoretical framework. Section 3 presents our experimental design. Section 4 empirically studies the economic and demographic determinants of underreaction to taxes, and section 5 quantifies the variance of this underreaction. Section 6 utilizes our theoretical framework to discuss the welfare implications of our empirical estimates. Section 7 concludes.

2 Theory

Here we present a simple model for analyzing the welfare impacts of taxation when consumers under- (or over-) react to not-fully-salient sales taxes. The theoretical results in this section directly

⁷For 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 (2013) for work related to income tax misperceptions. See Lusardi and Mitchell (2014) for a review of the many other domains of behavior in which financial literacy has been shown to play an important role.

⁸Results 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. Chetty et al. (2014) do not find that people’s savings decisions do not become less passive with respect to larger employer contributions. Hoopes et al. (2015), however, 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 (2014) show that consumers pay more attention to energy costs when gasoline prices are higher; Charness et al. (2010) show that study participants are less likely to exhibit the conjunction fallacy with higher incentives; but Zimmermann and Enke (2015) do not find that higher incentives reduce correlation neglect; and Brocas et al. (2014) do not find higher incentives increase subjects’ “depth of reasoning.” Also similar in spirit is the work by Andersen et al. (2011), who show that higher stakes lead to more profit-seeking behavior in ultimatum games.

motivate our empirical research questions about attention variation.

First, we show that in a market with no pre-existing taxes, the efficiency cost of introducing a small tax is increasing in the variance of mistakes. At the same time, we also show that aggregate demand responses, such as those analyzed in CLK, cannot identify this key statistic. This motivates our empirical research question about the variance of the magnitude of consumer mistakes.

Second, we generalize to a market with pre-existing taxes, and show that in this case it is also necessary to take into account how under- (or over-) reaction to taxes changes with the tax rate. The more salient higher taxes are, the higher the efficiency costs of increasing taxes. However, we show that this relationship cannot be inferred from locally estimated elasticities, motivating our investigation of how attention depends on the tax rate.

Third, we generalize to a richer welfare framework in which the policymaker has redistributive motives, and show that in this case it is also necessary to determine how underreaction to the sales tax covaries with household income. This shows that in addition to knowing how much variation there is in underreaction, it is also necessary to know *who* is most prone to underreacting to taxes.

2.1 Set-up

Consumers: There is a unit mass of consumers who have unit demand for a good x and spend their remaining income on a numeraire good y . 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, p the posted price of the product, and t the tax set by the policymaker.⁹

A fully optimizing consumer would choose $x = 1$ 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 $x = 1$ where $u(Z - p - \theta t) + v \geq u(Z)$, where θ —which may covary with v or be endogenous to t —denotes how much 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, $\theta = t/\hat{t}$.

⁹Note that we are assuming here that the policymaker is using a tax instrument with only one level of salience. See Goldin (Forthcoming) for a model (of otherwise optimizing consumers and no redistributive concerns) 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.

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). Because of this, 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 F generates smooth, downward-sloping demand curves, and that $\theta \geq 0$. With minor abuse of notation, we let $F(\theta|p, t)$ denote the distribution of θ for consumers who are indifferent between purchasing the product or not at posted price p and sales tax t . We define $E(\theta|p, t)$ and $Var(\theta|p, t)$ to be the mean and variance θ of consumers who are indifferent between purchasing the product or not at (p, t) .

Different theories of consumer mistakes have different implications for the shape of F . 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 $\theta > 1$. Moreover, the bias parameter θ may vary between individuals due to persistent individual differences like cognitive ability, or it may vary over time for a given individual due to changes in stimuli (see, e.g., DellaVigna and Pollet 2009).

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, and we let $\varepsilon_{D,p} = \frac{D_p(p,t)}{(p+t)/D(p,t)}$ and $\varepsilon_{D,t} = \frac{D_t(p,t)}{(p+t)/D(p,t)}$ denote the elasticities with respect to the posted price p and sales tax t , respectively.

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) .¹⁰ We relax the first assumption in Appendix B, following models such as those in Lockwood and Taubinsky (2015) and Farhi and Gabaix (2015).¹¹ Welfare analysis under these two assumptions 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.

Producers: We define production identically to CLK: price-taking firms use $c(S)$ units of the

¹⁰Assumption 2 implies that we leave out cognitive costs from our efficiency costs and welfare analysis. Although 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 should be negligible for a variety of models (Chetty et al., 2007).

¹¹Farhi and Gabaix (2015) also explore a general framework that doesn't rely on the second assumption.

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 $\varepsilon_{S,p} = \frac{\partial S / \partial p}{p/S(p)}$ denote the price elasticity of supply.

2.2 Efficiency Cost

We follow Auerbach (1985) in defining the efficiency cost of a tax for a heterogeneous economy by first defining excess burden for a given consumer and then aggregating. For a given consumer i , we define excess burden EB_i using equivalent variation. We let $x_i^*(p, t)$ denote consumer i 's choice of $x \in \{0, 1\}$ and we let $V_i(p, t, Z) = u(y - px^*(p, t, z) - tx^*(p, t, z)) + v_i x^*(p, t, z)$ 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. For a given consumer i , the excess burden of introducing a tax t in a previously untaxed market is $EB_i(t) = Z - e_i(p, 0, V_i(p, t, Z)) - R_i(t, Z)$. And the overall excess burden is defined to be $EB(t) = \int_i EB_i(t)$. Roughly, the excess burden is the loss in surplus accruing from consumers whose value for the product is above its price no longer purchasing it because of the tax.

We begin by studying the illustrative case in which there are no pre-existing taxes, producer prices are fixed, and u is linear (i.e., no income effects). Under these assumptions, the excess burden is simply $\int_{p \leq v \leq p + \theta t} (v - p) dF(v, \theta)$: it is the average value of the product, net of its price, to those consumers who no longer purchase it because of the tax. Analogous to Harberger (1964) and CLK, we derive a quadratic approximation, assuming that third and higher-order terms are negligible.

Proposition 1. *Suppose that producer prices are fixed ($\varepsilon_{S,p} = \infty$) and that there are no income effects (u is linear). For a smooth demand curve $D(p, t)$, and a small tax t , the dead-weight loss corresponding to that tax is approximately*

$$\begin{aligned} EB(t) &\approx \frac{1}{2} t^2 [E(\theta|p, t)^2 + Var(\theta|p, t)] D_p(p, t) \\ &= \frac{1}{2} t^2 [E(\theta|p, t)^2 + Var(\theta|p, t)] D(p, t) \frac{\varepsilon_{D,p}}{p+t} \end{aligned} \quad (1)$$

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

Proposition 1 provides a general formula for the 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 $E(\theta) = 0.25$. When all consumers are homogeneous and have a $\theta = 0.25$, equation (2) shows that the excess burden is $(0.25)\frac{1}{2}t^2D(p, t)\frac{\varepsilon_{D,t}}{p+t}$; 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 $\theta = 0$. In this case, we still have $E(\theta) = 0.25$, but (2) implies that the excess burden is approximately $\frac{1}{2}t^2D(p, t)\frac{\varepsilon_{D,t}}{p+t}$, which is exactly identical to the inference that would be made by an analyst assuming that consumers optimize perfectly, and using the tax elasticity of demand.

Figure 1 provides a graphical representation of the formula. The green line represents demand as a function of t , keeping the initial price p_0 fixed. The blue line represents what demand would look like if all consumers reacted to the tax correctly. An analyst assuming perfect optimization and observing only the green demand curve would conclude that the excess burden is given by a triangle under the green line. An analyst who knows the average θ , but assumes that it is homogeneous, would conclude that the excess burden is given by the area of the red triangle. The red triangle corresponds to the excess burden calculation given in CLK. However, as Proposition 1 shows, the more heterogeneous consumers are, the greater the excess burden. This additional distortion is captured by the orange triangle, the height of which is given by $t\frac{Var(\theta|p,t)}{E(\theta|p,t)}$.

In fact, equation (2) shows that even if consumers underreact to the tax on average ($E(\theta) < 1$), the excess burden could still be greater than the inference that would be made under the assumption of perfect optimization. This occurs when the variance is sufficiently large due to some consumers overreacting to the tax ($\theta > 1$). Similarly, equation (1) shows that even when consumers underreact to the tax on average, a sufficiently large variance can make the excess burden greater than what it would be if all consumers reacted to the tax fully (or if posted prices were tax-inclusive).

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.¹² This intuition comes out most starkly when supply is not perfectly elastic, a case which we treat below.

¹²Glaeser and Luttmer (2003) study a neoclassical analog to this departure from traditional deadweight loss analysis: they 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.

2.2.1 Endogenous Producer Prices

We now generalize Proposition 1 to the case in which producer prices are endogenous. In this case, assume that firm profits are paid back to the consumers using the numeraire y . Excess burden is now given by $EB_i(t) = Z - e(p, 0, V_i(p, t, Z)) - R(t, Z) + \pi_0 - \pi_1$, where $\pi_0 - \pi_1$ is the change in producer profits. We let $p(t)$ denote the equilibrium price as a function of t .

Proposition 2. *Suppose that utility is quasilinear ($u(y) = y$). Then the excess burden of introducing a small tax t in a previously untaxed market is approximately*

$$EB(t) \approx -\frac{1}{2}t^2 \left[E(\theta|p, t) \frac{d}{dt} D(p(t), t) + Var(\theta|p, t) D_p(p(t), t) \right]$$

where the total derivative $\frac{d}{dt} D(p(t), t)$ is evaluated at $(p(t), t)$, and $D_p(p, t)$ denotes the partial derivative of demand with respect to price evaluated at $(p(t), t)$.

When $Var(\theta|p, t) = 0$, Proposition 2 shows that the excess burden calculation is nearly identical to the case of fixed producer prices, except with the partial derivative of demand replaced with the total derivative of the equilibrium quantity purchased. In contrast, when there is variation in θ , excess burden depends on both the total derivative and the partial derivative of demand. This is because the extent to which dispersion in θ generates misallocation depends on how sensitive consumer choices are to differences in the perceived final price, $p + \theta t$, of the product. A particularly stark and illustrative case arises when supply is inelastic:

Corollary 1. *Suppose that supply is inelastic ($\varepsilon_{S,p} = 0$). Then with quasilinear utility, the excess burden of introducing a small tax t in a previously untaxed market is approximately*

$$EB(t) \approx -\frac{1}{2}t^2 Var(\theta|p, t) D_p(p, t)$$

Corollary 1 shows that when supply is inelastic—and thus the equilibrium quantity produced by the market does not change—the efficiency cost of a small tax t depends only on the variance of bias and the change in demand. This is because when the equilibrium quantity does not change, all of the efficiency cost is generated by misallocation, the extent of which is proportional to the variance of θ . Thus when supply is inelastic and consumers are heterogeneous in θ , excess burden would be *smaller* if the tax was made perfectly salient to all consumers.

Corollary 2. *Suppose that $Var(\theta|p, t) > 0$. Then there exists a small enough $\varepsilon_{S,p}/\varepsilon_{D,p}$ such that excess burden would be smaller if the tax was perfectly salient to all consumers.*

Building on the stark result in Corollary 1, Corollary 2 shows more generally that in the presence of heterogeneity, whether making a tax less salient is good or bad depends on how elastic supply is relative to demand. When producers bear the incidence of the tax, the total change in quantity

demand, $\frac{d}{dt}D$, will be small, and thus efficiency costs will be small in the presence of salient taxes. However, excess burden may still be large with non salient taxes, because the efficiency cost of mis-sorting, which is proportional to $Var(\theta)D_p$, may still be substantial.

2.2.2 Pre-existing Taxes and Endogenous Underreaction

We now study a case in which there are moderate pre-existing taxes, such as those in the US, but attention increases by a non-negligible amount when the taxes are increased to, e.g., triple their value. Even though a tripling of US sales taxes would correspond to an average increase of only 14 percentage points, our experimental results show that this can have a substantial impact on attention.

Proposition 3. *Suppose that u is linear and that the pre-existing tax t is of the same order as Δt . Then the excess burden of a small increase Δt on top of the pre-existing tax t is*

$$EB(t + \Delta t) - EB(t) \approx \left(t\Delta t + \frac{\Delta t^2}{2} \right) E[\theta^2|p, t + \Delta t]D_p \quad (3)$$

$$+ \left(\frac{t^2}{2} \right) (E[\theta^2|p, t + \Delta t] - E[\theta^2|p, t]) D_p \quad (4)$$

The main insight of Proposition 3 is that when increasing a pre-existing tax has a non-negligible impact—given by $\Delta E[\theta^2] = E[\theta^2|p, t + \Delta t] - E[\theta^2|p, t]$ —on consumer bias, it is necessary to account for that in the computation of excess burden. Intuitively, increasing a tax increases excess burden through two channels. First, raising the tax decreases demand for x holding underreaction constant—this effect corresponds to the right-hand side term in line (3) of the formula. Second, increasing the tax may decrease underreaction. This decrease in underreaction in turn creates additional excess burden by making consumers more attentive to the tax, and thus less likely to purchase the product—this is captured by the term in line (4).¹³

2.2.3 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 principals as the baseline excess burden formula without income effects, and thus we relegate this to Appendix A.2. In the appendix, we show that with income effects there is an additional efficiency when consumers underreact to taxes, because they consume too much y relative to x . Thus even when $E[\theta] = 0$ and $D_t = 0$, there can still be deadweight loss.

¹³In a representative consumer model, Reck (2014) shows that if attention is “sufficiently endogenous” in the sense that consumers are fully debiased for a high enough tax, then there must exist a sales tax t such that the marginal impact on excess burden from increasing t is higher than the marginal impact on excess burden from increasing a salient tax.

2.3 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.¹⁴

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 . 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_ω . 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 4. *Set $\bar{g} := \int g_\omega$. For a small tax t ,*

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

Proposition 4 shows that just as excess burden is decreasing 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 1 when $g_\omega = \lambda = 1$ for all ω , Proposition 4 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

¹⁴We 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).

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.

2.4 Extensions and Optimal Tax Implications

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.

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 .

2.5 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 degree of underreaction to taxes relative to prices, $D_t(p, t)/D_p(p, t)$.

We provide more general results about the conditions under which excess burden can be inferred from $D(p, t)$.¹⁵ First, we ask under what conditions knowledge of the whole demand curve D can

¹⁵We focus here on the question of when excess burden can be identified, because the conditions under which welfare with redistributive concerns is identified are the same as the conditions under which excess burden is identified, combined with the requirement that it is possible to estimate the covariance between bias and welfare weights.

identify efficiency costs. Second, we ask under what conditions local knowledge of D is sufficient to identify the efficiency costs of a small change in the tax. The question about local knowledge corresponds to the typical implementation of the sufficient statistics approach (Chetty, 2009b) that relies only on locally estimated elasticities.

In Appendix A.1 we prove two main results. 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 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, t , then $D(p, t)$ can never identify the dispersion, and thus welfare. This is particularly intuitive for efficiency costs arising from a small tax t . In this case, the average θ is identified by D_t/D_p , but 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 common 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, and has also been previously used by Benjamin et al. (2014) as well as in our own work reported in Rees-Jones and Taubinsky (2015). 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 2 provides a synopsis of the experimental design. The design had four parts: 1) 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 receive a \$20 shopping budget, 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

Because the main goal of this analysis is to illustrate the difficulty in obtaining welfare estimates from aggregate demand data with heterogeneous biases, we focus here on the simplest case of identifying excess burden.

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,” which we sometimes denote by $C=0x$ for shorthand 2) the “Standard Tax Arm,” which we sometimes denote by $C=1x$ for shorthand and 3) the “Triple Tax Arm,” which we sometimes denote by $C=3x$ for shorthand. 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.¹⁶

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.¹⁷

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 3 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 no tax, standard tax, and triple tax decision environments In the *no tax* decisions consumers were told that “In contrast to what shopping is like at your local store, no sales tax will

¹⁶An 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 deconvolute 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.

¹⁷Local tax rate data is drawn from the April, 2015 update of the “zip2tax” tax calculator.

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 2, 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.

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.

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 the 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 3.

Incentive compatibility Decisions in the experiment were incentive compatible. Study participants had a 1/3 chance of being selected to receive a \$20 budget. Participants were informed of this prior to making any decisions, but they did not know if they received the budget or not 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. For the decisions that were selected, 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 corresponding the 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. The possible answers were 0, the tag price, the tag price plus the standard sales tax, and the tag price plus triple the sales tax. 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 G 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 Economic and Demographic Determinants of Underreaction to Taxes

4.1 Sample Selection, Demographics, and Balance

4.1.1 Sample

A total of 4,329 consumers completed the experiment. For this sample, 3,068 correctly answered instruction-comprehension questions in both module 1 and module 2. For our primary analysis, we exclude those who did not correctly answer both comprehension questions correctly. Out of the remaining 3068 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. Out of the remaining 3038 consumers, 10 consumers reported living in a state with no sales tax. We exclude these consumers as well, because our recruitment criteria for ClearVoice specifically asked to select panel members only from states with

a positive sales tax.¹⁸

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 3000.

Unsurprisingly, the 29% of consumers who did not pass the comprehension questions do not react to the differences in taxes in our conditions. We exclude these consumers from our main sample, because the misoptimization these consumers exhibit is due to not (carefully) reading the instructions. This is a mistake that is at best imperfectly correlated with how they would respond to taxes if they were to read the instructions, and is thus a potentially serious confound for the questions about underreaction to taxes that our study is about. We discuss this in more detail in Section 4.7, and we address any potential confounds created by selection on comprehension in two ways. First, in Appendix C, we provide econometric results and bounding approaches for guaranteeing the robustness of our results to selection on comprehension. Second, in Appendix E.9, we show that all of our key results are robust to the using the full sample.¹⁹

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 40% college-educated—is similar to the US population on these basic demographics. There are no significant differences in demographics between Arm 1 vs. Arm 2 (F -test $p = 0.38$), Arm 2 vs. Arm 3 (F -test $p = 0.36$) or Arm 1 vs. Arm 3 (F -test $p = 0.44$), and as the table also shows.

There are, however, small but statistically significant differences in the likelihood that consumers pass the comprehension questions in the different arms of the study.²⁰ 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,

¹⁸These 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.

¹⁹We also included questions to check if participants understood the BDM. 78% of participants passed those comprehension questions, and we show in Appendix E.11 that our results are robust to restricting to this sample. We are far less concerned about potential misunderstanding of the BDM 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 systematic misunderstanding do not confound estimates of θ (only estimates of true valuations): even if participants scale their answers up or down because of systematic misperceptions of the mechanism, the ratio of the bids will still be $1 + \theta\tau$ (in expectation).

²⁰The 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 of equal pass rates is rejected for any pair of arms at the 5% significance level.

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. As we already noted, we will show in Section 4.7, as well as in appendices C and E.9, 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 4 provides a summary of the demand curves as functions of before-tax price. To construct the figure, we start with demand curves $D_k^{C,m}(p)$ 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* $D_{avg}^{C,m}(p) := \frac{1}{20} \sum_k D_k^{C,m}(p)$ 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 4 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 4 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 ($p = 0.734$), between the no tax and triple tax arms ($p = 0.302$), and between the standard tax and triple tax arms ($p = 0.500$). In Appendix E.1, we also show OLS and quantile regressions comparing the average willingness to pay in module 2, which similarly detect no differences.²¹

While consumers react to taxes, they do not react to taxes as much as perfect optimization would imply, as we show in figure 5. Panel (a) of figure 5 is identical to panel (a) of figure 4, 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 5 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.

Consumers’ valuations of the products reflect market prices. The average Amazon.com price for the products was approximately \$10,²² while consumers’ average valuation for the products was approximately \$6. An OLS regression (clustering standard errors by subject) of module 2 valuations

²¹The standard and triple arms have average tag prices that are, respectively, 8.7 cents ($p = 0.392$) and 3 cents ($p = 0.765$) lower than the average tag prices in the no tax arm. We thus find no evidence for anchoring or demand effects.

²²Prices were recorded by our research assistant in February 2015. See Appendix G for the products, prices, and descriptions.

on Amazon prices obtains a regression coefficient of 0.36 (t -stat = 82).²³

4.3 Econometric Framework for Quantifying (Differences in) Underreaction

We now present our baseline econometric framework for studying how underreaction to taxes varies by experimental condition and by observable demographics—which we employ in Sections 4.4 and 4.6. 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 depend on the order in which product k appears in the experiment—what we will refer to as *order effects*—and even after controlling for order, it is only a noisy estimate of θ_{ik} because study participants’ reported values for the product fluctuate. Let Γ_{ik} denote a 40×1 vector of decision order dummies that contains a 1 in the entries corresponding to the decision numbers (1-40) in which product k appears for participant i . Let X denote the vector of covariates of θ , and let $Z = (\Gamma, X)$. We assume that order effects do not vary by experimental arm in our final sample, which allows us to identify order effects from behavior in the “no tax” treatment arm:

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

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

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

The model above implies the following moment conditions:

$$\begin{aligned} E[Z'_{ik}y_{ik}] &= Z'_{ik}\beta Z_{ik} \quad \text{if C=0x} & (5) \\ E\left[X'_{ik}\left(\frac{y_{ik} - \beta Z_{ik}}{\tau_i}\right)\right] &= X'_{ik}\alpha X_{ik} \quad \text{if C=1x or C=3x} & (6) \end{aligned}$$

Equation (5) identifies any order effects in the data using the no tax arm. These order effects are partialled 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 the data, we will assume that the coefficient on a demographic covariate such as, e.g., an income

²³Given that Amazon.com prices fluctuate over time and across consumers, and that consumers on average valued the products at 60% of their price, the 0.36 coefficient is substantial. Overall, consumer valuations were typically below the Amazon.com prices. Consumers were willing to pay more than the Amazon.com price only 13% of the time.

quartile, is the same in each arm, but use a separate moment equation (6) for each arm (i.e., assume (6) holds independently in both the standard and triple tax arm). In our pooled analysis, the system is thus over-identified, and we use the two-step GMM estimator to obtain an approximation to the efficient weighting matrix.

Three notes are in order about our estimation procedure. *First*, using OLS to simply estimate the linear model (7) below would not yield consistent estimates of average θ .

$$E[y|X, \Gamma] = \alpha_1\tau + \alpha_2\tau X + \beta_1X + \beta_2\Gamma. \quad (7)$$

Suppose that we are interested in calculating the average parameter θ across all study participants. Because the linear model is misspecified when underreaction is endogenous to the tax rate, an estimate $\hat{\alpha}_1$ obtained from equation (7) 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.²⁴ While these endogeneity concerns lead us to prefer the GMM approach as a primary specification, we note that all of our qualitative results about demographic and economic covariates are also obtained when estimating (7) using OLS, as shown in Appendix E.10.

Second, 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.

Third, note 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.

4.4 Average Underreaction to Taxes by Experimental Arm

Table 2 presents our estimates of average θ in each arm, approximated by $E\left[\frac{\log(1+\theta\tau)}{\tau}\right]$, 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

²⁴To take a concrete illustration from the next section, 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 (7) 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.

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 $p_2^{ik} \geq 5$. In section 4.7 we show that the results are robust to worst-case assumptions about which subjects get excluded for not passing the comprehension questions.

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 over-react to the triple tax if they are surprised by the unusual scenario in which taxes are three times the size of what they usually are (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. As we show in Table A2 in Appendix E.2, however, we are underpowered for such an analysis, because there is too little variation in state tax rates—we can neither reject the null hypothesis of no effect nor the effect sizes in table 2.²⁵

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 . In the context of our experiment, the question we ask is whether θ is higher when the module 2 valuation is higher. We operationalize this by dividing all consumers (from all three arms) into three bins corresponding to $p_2^{ik} < 5$, $p_2^{ik} \in [5, 10)$, 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

²⁵As table 1 shows, the average tax rate in our experiment is 7.32%, with a standard deviation of only 1.15%, and a maximum tax rate of 10.6%.

we also cannot reject a modest 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 $p\tau$, but it does not rule out that our estimates are driven by the possibility that consumers who are willing to pay the most for the products are also the consumers who have the highest θ . Whether this is a “confound” depends on the question of interest: for welfare computations such as those in Section 2.2, for example, it does not matter which of these mechanisms is driving the relationship between average θ and price. Columns (4)-(6), however, report a more robust test of whether consumers are more attentive to taxes on more expensive products. The model estimated in these columns controls for person-specific fixed effects in θ in measuring how θ varies with module 2 price p_2 . The test thus identifies only off of how θ varies by p_2 bin *within* each consumer. Appendix E.3 formally documents how we modify the strategy in Section 4.3 to implement this estimation. While estimates are slightly attenuated toward zero as compared to the estimates in the first three columns, we again see the similar pattern that on average θ appears to be somewhat smaller when $p_2^{ik} < 5$ than when $p_2^{ik} \geq 5$.

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.”²⁶

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. 51% of consumers know their tax rate exactly, about 73% are within 0.5 percentage points, and about 85% are within 1 percentage point. The average of beliefs is 7.32%, while the average actual tax rate of consumers in the study is 7.22%, indicating almost no mean bias. The lack of any significant mean bias shows that incorrect beliefs cannot explain the average underreaction we see in our data. To provide a graphical summary of how perceived beliefs vary with the actual tax rate, we construct Figure 8 in Appendix E.8, which shows a near-perfect 45-degree line relationship between perceived and actual tax rates. We

²⁶If 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.

conclude that incorrect beliefs are a negligible source of consumer mistakes on average, consistent with CLK’s survey results from consumers in a California store.

4.6.2 Numeracy, Financial Literacy and Income

We now demonstrate that the ability to compute taxes, financial literacy, and household income are all predictive of underreaction to taxes. We first describe these three key covariates, and then present the results.

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, 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.

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.²⁷ For completeness, we include these questions in Appendix F. 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.²⁸

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.²⁹

To provide a graphical summary of how θ varies by income, we construct Figure 6. The figure plots estimates of average θ by income deciles, and also provides a local polynomial estimate of the relationship between θ and income. Particularly for the triple tax arm, where we have the most statistical power, the figure shows that average θ is steeply increasing with household income.

²⁷Previous work has shown that financial literacy is associated with mistakes in other domains, including incurring overdraft fees (Stango and Zimman, 2014), incorrectly valuing annuities (Brown et al., 2013), and not saving enough for retirement (Lusardi and Mitchell, 2007a,b).

²⁸Our measure of tax numeracy and financial sophistication are also correlated. Financially sophisticated consumers have a 12 percentage point greater likelihood of correctly answering the tax computation question ($p < 0.01$).

²⁹According to the 2010 census, the quartile thresholds are 25k, 50k, 90k.

Table 4 presents the results for all three of our covariates, pooling data across the standard and triple tax arms for power. Column (1) of the table shows that consumers who were able to correctly compute the tax have a significantly higher θ , on average. This result suggests a bounded-rationality model in which people ignore attributes of a financial decision that they do not fully understand.

Column (2) of table 4 shows that financially sophisticated consumers underreact less to the sales tax. This suggests that incorporating sales taxes into one’s decision requires financial skill, and thus that consumers who are more financially literate are more likely to correctly react to the sales tax.

Column (3) shows that consumers in the first quartile of the income distribution underreact to taxes twice as much as consumers in the fourth quartile of the income distribution. This difference is significant at $p < 0.01$ in both the pooled and triple tax samples. The table shows that consumers in the third quartile also attend more to taxes than consumers in the first quartile ($p < 0.05$), though we do not detect a difference between the first and second quartile.³⁰

Column (4) shows that when we include all three covariates, all of the coefficients are dampened, but they remain significant at at least the 10% significance level. The coefficients are dampened because all three of the variables are strongly correlated with each other, suggesting that at least some of the relationship between θ and income is likely due to the fact that higher-income consumers are more financially sophisticated and numerate. Indeed, Figure 7, which plots the ability to compute taxes and financial literacy by income deciles, shows that financial sophistication and the ability to compute taxes increase dramatically with income.

In appendix E.5, we replicate Table 4 with the inclusion of demographic controls such gender, age, educational attainment, and race. The three key covariates studied in this section remain significant. Of all the other demographic covariates, only age is significantly associated with θ , with older people underreacting more to the sales taxes. We also replicate Table 4 for each experimental arm in Appendix E.7.

4.7 Robustness to Selection on Comprehension Questions

A limitation of any experiment other than a natural field experiment is the possibility that study participants don’t fully understand the experimental environment.³¹ In our context, if some study participants behave as if there is a sales tax even though they are in an experimental condition without a sales tax, while other study participants behave as if there is no sales tax even though they are in an experimental condition with a sales tax, then that amplifies our estimates of how much

³⁰We can also estimate an elasticity of attention θ by income z : $\varepsilon_{\theta,z} := \frac{dE[\theta|z]}{dz} \frac{z}{\theta}$. Veiga and Weyl (2015) study a monopolist offering products with shrouded attributes to heterogeneously misoptimizing consumers, and show that this elasticity is the key statistic for determining the monopolists’ incentive to increase or decrease the shrouded add-on price. To estimate an average elasticity, we estimate the model $\log(\bar{\theta}) = \bar{\theta}_0 + \varepsilon_{\theta,z} \log(z)$, or $\bar{\theta} = e^{\bar{\theta}_0 + \varepsilon_{\theta,z} \log(z)}$. We estimate this using the method of moments, controlling for order effects as before, and excluding the extreme income observations below \$1000 and above \$250,000. In the pooled date, we estimate $\hat{\varepsilon}_{\theta,z} = 0.37$, with a standard error of 0.12. Although it is hard to extrapolate from our setting to other settings such as, e.g., overdraft fees, our results nevertheless provide an initial data point for calibrations such as those in Veiga and Weyl (2015).

³¹To be more specific, we mean that participants don’t read the instructions, or don’t fully understand the parameters that are specific to the particular experimental environment, but don’t map on to natural settings.

study participants underreact to sales taxes. Perhaps more importantly, confusion is a confound of any analysis of individual differences: for any potential covariate of underreaction, it is important to ensure that it is indeed a covariate of underreaction, rather than a covariate of the propensity to read instructions carefully.³² For this reason, our final sample includes only study participants who correctly answered the tax comprehension questions in both module 1 and module 2 of the experiment.

Selecting on comprehension, however, may introduce differences in the final samples across the three arms. While our samples do not appear to 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. However, we show in Appendix C that under mild assumptions, 1) the possibility of selection on unobservables does not bias our analysis of heterogeneity and 2) we provide a lower bound for the difference between average θ in the triple and standard tax arms.

Using the lower bound in part 3 of Proposition 10 in Appendix C, we address the concern that some of the difference in average θ may be due to the final samples in the standard and triple tax arms not being comparable because of the different rates at which consumers passed the comprehension questions. 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 ($p = 0.085$) when using all data, at the 5% significance level ($p = 0.041$) when conditioning on module 2 price $p_2^{ik} \geq 1$, and at the 1% significance level ($p = 0.005$) when conditioning on $p_2^{ik} \geq 5$.

4.8 Comparison to Related Studies

While our study is unique in its theoretical framework and the experiment designed to estimate the sufficient statistic formulas we derive, several existing studies have addressed elements of our empirical investigation. Feldman, Goldin and Homonoff (2015, henceforth FGH) run a 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 demographic 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

³²These concerns about comprehension are of course a concern not just for our experiment, but for any experiment studying individual differences. For example, any experiment studying gender differences must ensure that differences in behavior between men and women are due to actual economically relevant differences, and not differences arising from the possibility that men and women have different levels of comprehending experimental instructions.

interval, the significantly greater power of our design allows us to reject the null hypothesis that the ratio equals 1.³³

Interestingly, our result about the covariance of θ and income contrasts with Goldin and Homonoff (2013), who find that low-income cigarette smokers *over-react* by a factor of four to the sales taxes (specifically, they find that they are four times more elastic to sales taxes than excise taxes), and that this overreaction is significantly smaller for higher-income smokers. It is hard to directly compare these results to ours, because the relationship between θ and income may be different for smokers than for the general population, and it may also be different for cigarettes than for other goods if consumers are particularly attentive to cigarette taxes.³⁴ On the other hand, the ClearVoice panel may also not be representative of the general population. In our data, we do not have enough power to estimate a relationship between θ and income for the smoker subsample³⁵ (though we can reject that average θ is greater than 1 for each income quartile of the smoker subsample).³⁶

5 The Variance of Consumer Underreaction

Having established that underreaction varies across both economic and demographic conditions, we now return to our first theoretical question: the extent of individual differences in θ . We present our methodology in Section 5.1 and implement an estimate of the lower bound in Section 5.2. We discuss the advantages of our approach over other possible approaches in Section 5.3.

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

As the results in Section 2.2 show, the statistic needed for welfare analysis is $Var[\theta|p, \tau]$ —the variance of θ of consumers who are indifferent between buying the product or not at a given posted price p and tax rate τ . The statistic we estimate is thus $E_{p_1, \tau}[Var[\theta|p_1, \tau]]$; that is, the average, over all pairs (p_1, τ) , of the variance of θ of consumers who are marginal at each price-tax pair (p_1, τ) . 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 . 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 .

³³Controlling for subject comprehension is crucial, however, because confusion creates unresponsiveness to stakes and thus attenuates the ratio; we would find the ratio to be only 1.26 if we did not exclude study participants who did not correctly answer the comprehension questions. FGH do not have an explicit test for comprehension. In our data, while most participants report that the experimental instructions were very clear, many still do not correctly answer the comprehension questions.

³⁴A second potential reason for the different results is that identification strategies differ. If it is the case that low-income consumers tend to spend a higher fraction of their income on tax-exempt goods than high-income consumers, then comparing the excise tax vs. sales tax elasticity ratio for low- vs. high-income consumers can yield biased estimates.

³⁵Just over 20% identify as smoking cigarettes, and just under 30% identify as smokers more generally

³⁶We do have enough power, however, to confirm that financial literacy is increasing with income in the smoker subsample. To the extent that financial literacy is a covariate of θ and that carries over to the smoker subsample, this would suggest that θ is increasing in income in the smoker subsample as well.

Roughly, the idea behind our lower bound is to use survey question responses to partition study participants into subgroups with different average θ 's, and to then compute the variance of the subgroup means. The approach can be applied to any survey question or observable characteristic that is correlated with θ . In our empirical implementation, however, we will rely on 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. The self-classifying survey question, which we describe in more detail in Section 5.2, asks study participants in the standard and triple tax arms how they would behave if there were no taxes in those conditions.

Formally, 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 $R = L$. Our technique can be immediately generalized to any observable characteristic R that can take on any number of finite values. We set $\phi := \frac{\log(1+\theta\tau)}{\tau}$, $\mu(p, \tau) := E[\phi|p, \tau]$, $\bar{\phi}(r, p, \tau) := E[\phi|r, p, \tau]$. For short-hand, we let $\bar{\theta}_r := E[\bar{\phi}|R = r]$; that is, $\bar{\theta}_r$ is the (approximately) average θ of consumers with survey response $R = r$.

Proposition 5.

$$E_{p_1, \tau}[Var[\theta|p_1, \tau]] \geq E[Var[\bar{\phi}|\tau, p_1]] \tag{8}$$

$$\geq Pr(R = H) (E[\bar{\phi}|R = H] - E(\mu|R = H))^2 \tag{9}$$

$$+ Pr(R = M) (E[\bar{\phi}|R = M] - E(\mu|R = M))^2 \tag{10}$$

$$+ Pr(R = L) (E[\bar{\phi}|R = L] - E(\mu|R = L))^2 \tag{11}$$

Proposition 5 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 $Var[\theta_R]$ 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) .

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) , call it $\mu(p_1, \tau)$ and then 2) compute the average μ with respect to the 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 $E_{p_1, \tau}[Var[\theta|p_1, \tau]] \geq E\left[Var\left[\frac{\log(1+\theta\tau)}{\tau}|\tau, p_1\right]\right]$, 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 $E_{p_1, \tau}[Var[\theta|p_1, \tau]] \geq E[Var[\bar{\phi}|\tau, p_1]]$. Third, we arrive at the final quantity in (9)-(11) through an application of the Cauchy-Schwarz inequality. In Section 5.3 we explain why the statistic in (9)-(11) is econometrically a significantly more tractable statistic to estimate than the statistic in (8).

5.2 A Lower Bound for the Variance of Mistakes: Estimation

5.2.1 The Survey Instrument

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 A4 in the appendix 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 $p < 0.01$).³⁷

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 cannot estimate any order effects associated with the survey responses. We thus make the additional assumption A2 that if survey responses R are predictive of behavior, it is solely because they are correlated with θ :

A2 $E[y_{ik}|\theta_{ik}, \Gamma_{ik}, R] = E[y_{ik}|\theta_{ik}, \Gamma_{ik}]$, where Γ_{ik} is the vector of order dummies

A2 implies that for both arms,

$$E\left[\frac{y_{ik} - \beta\Gamma_{ik}}{\tau_i} | R = r\right] = E\left[\frac{\log(1 + \theta_{ik}\tau_i)}{\tau_i} | R = r\right], \quad (12)$$

where as before β is identified from the no tax arm. 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,”

³⁷However, 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.

is in the neighborhood of 0.5 for consumers who answer “Maybe a little”, and is in a neighborhood 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 estimate the lower bound in Proposition 5.

5.2.2 Lower Bound Estimation

A challenge in estimating the lower bound from Proposition 5 is estimating the terms $E(\mu|R = r)$. Because our dataset is finite, we cannot obtain an accurate 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 $\{\mathbf{p}_j\}_{j=1}^{15}$ denote the fifteen cells $[0, 1], [1, 2], \dots, [14, \infty)$ and let $\{\boldsymbol{\tau}_j\}_{j=1}^5$ denote the five cells $(0, 6\%], [6\%, 7\%], \dots, [9\%, \infty)$. 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

$$\tilde{\mu}(p_1, \tau) = E \left[\frac{\log(1 + \theta_{ik}\tau_i)}{\tau_i} | p_1^{ik} \in \mathbf{p}(p_1), \tau_i \in \boldsymbol{\tau}(\tau) \right]. \quad (13)$$

As the cell sizes converge to zero, $\tilde{\mu}$ will converge to μ . The statistic with which we approximate the lower bound from Proposition 5 is now

$$\sum_{r \in \{L, M, H\}} Pr(R = r) (\bar{\theta}_r - E[\tilde{\mu}|R = r])^2 \quad (14)$$

To estimate (14), we estimate each $\bar{\theta}_r$ using the empirical moment version of the left-hand-side of (12). We estimate $\tilde{\mu}(p_1, \tau)$ using the empirical moment counterpart of

$$E \left[\frac{y_{ik} - E[y_{ik}|C = 0\mathbf{x}, p_1^{ik} \in \mathbf{p}(p_1), \tau_i \in \boldsymbol{\tau}(\tau)]}{\tau_i} | p_1^{ik} \in \mathbf{p}(p_1), \tau_i \in \boldsymbol{\tau}(\tau) \right] \quad (15)$$

where $E[y_{ik}|C = 0\mathbf{x}]$ 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. See Appendix E.4 for further details.

Table 6 presents the results. The top row, in bold, displays our estimates of (14) for both the standard 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 0.124.³⁸ 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 produce substantial implications for the efficiency costs of taxation, as we show in Section 6.1.

5.3 Discussion of our Approach

Our approach vs. “direct” approaches for estimating the variance A common summary statistic presented in experiments studying structural parameters is to provide a summary of the distribution of individual-level estimates of the parameters. However, because each point estimate is only a noisy measure of the true parameter, as the number of participants increase, the distribution of point estimates approximates the convolution of the distribution of true parameters and noise. In our experiment, note that while $\nu_{ik} := 1 - p_2^{ik}/(\tau_i p_1^{ik})$ provides a noisy estimate of θ_{ik} , the variance of ν_{ik} vastly overestimates the variance of θ_{ik} because of the randomness in study participants’ decision-making that this variable picks up.

This approach could be modified, however, by first estimating the variance of noise using the no tax arm of the experiment, and then subtracting that variance from the variance of ν_{ik} estimated in the standard and triple tax arms. Theoretically, this approach provides unbiased estimates of the variance of θ_{ik} , and can be extended to estimate higher order moments. However, while our experiment is well powered to estimate first moments, our data is unfortunately too noisy to provide well-powered estimates of higher order moments.³⁹

³⁸The source of the bias is that any noise in our estimates of $\bar{\theta}_r$ or $E[\mu|R = r]$ amplifies the statistic in (14) 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} .

³⁹An additional 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.

Other possible lower bounds Estimating the seemingly simpler and tighter bound in (8) is far less econometrically tractable. Doing so would require estimating the variance of $\mu(p, t)$ within each of 75 cells $\{\mathbf{p}_j \times \boldsymbol{\tau}_j\}$, and then taking the average of those variances. This computation involves 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 two random variables is smaller than the average of the variance of those two random variables.⁴⁰

As a final note, we reiterate that the methodology used in this section is not limited to the self-classifying survey question. For example, a similar kind of analysis could be performed by sorting consumers into groups based on financial literacy or tax numeracy. The reason we focus on the self-sorting survey question is because it is by far the most predictive, and expanding the number of categories would reduce statistical power.⁴¹

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. The benchmark that we consider throughout is inferences assuming exogenous and homogeneous θ , and we translate all of our empirical estimates into how much they change welfare inferences relative to this benchmark. We summarize the results from this section in Table 7.

6.1 Individual Differences

To translate the estimates from Section 5.2 into excess burden estimates, we use the formula in Proposition 1, which expresses excess burden in terms of the mean and variance of θ . We consider a case in which there are no income effects and supply is perfectly elastic, because as shown in Proposition 2, the relative importance of individual differences increases as the elasticity of supply decreases.⁴² For the illustrative calculations here, we assume that for the price p and tax t in question, 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 $E_{p,t}[Var[\theta|p, t]]$.

⁴⁰To expand on the intuition, consider i.i.d. nuisance parameters $\epsilon_1, \dots, \epsilon_n$. Now consider the following two statistics: 1) $\frac{1}{n} \sum \epsilon_i^2$ and 2) $(\frac{1}{n} \sum \epsilon_i)^2$. The first statistic has mean and variance given by $\frac{1}{n} E[\epsilon_i^2]$ and $\frac{1}{n} Var[\epsilon_i^2]$, respectively, while the second statistic has mean and variance given by $\frac{1}{n^2} E[\epsilon_i^2]$ and $\frac{1}{n^2} Var[\epsilon_i^2]$, respectively. Thus, the second statistic has an order of magnitude smaller mean and variance. This difference between the two statistics is essentially the reason for why the second bound we derive can be estimated with tighter standard errors and a smaller mean bias.

⁴¹To obtain intuition for this, note that if we had as many bins as consumers in our experiment, then our estimates of the lower-bound would mostly be reflecting the noise in consumers' decisions, and thus would be very imprecise and biased. Increasing the number of bins increases the lower bound in expectation, but it also lowers statistical precision and increases bias.

⁴²And as discussed in Section 2.2.3 and further in Appendix A.2, income effects exacerbate excess burden, with that additional effect also increasing in the variance of the bias.

Let $EB_{neoclassical}$ 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.⁴³ 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 .⁴⁴ 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 equation (2) in Proposition 1, the actual excess burden is $EB \geq (0.25 + 0.124/0.25)EB_{neoclassical} = (0.75)EB_{neoclassical}$. Using the 5% confidence bound for the lower-bound statistic, we have $EB \geq (0.45)EB_{neoclassical}$.

For the triple tax arm, $EB_{homogeneous} \approx (0.48)EB_{neoclassical}$. However, by equation (2) in Proposition 1, the actual excess burden is $EB \geq (0.48 + 0.094/0.48)EB_{neoclassical} = (0.68)EB_{neoclassical}$. Using the 5% confidence bound for the lower-bound statistic, we have $EB \geq (0.61)EB_{neoclassical}$.

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. Moreover, the 5% confidence bounds produce meaningful estimates as well. We stress that these estimates are lower bounds, and that the actual impact of individual differences is likely to be much greater.

6.2 Endogenous Attention

Continuing with the excess burden formulas in Section 2.2, we now turn to the implications of endogenous attention that we formalize in Proposition 3. 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 3 implies that

$$EB(t + \Delta t) - EB(t) \approx (t\Delta t + (\Delta t)^2/2)(0.5)^2 D_p + \frac{t^2}{2}(0.5^2 - 0.25^2) \approx 0.84t^2$$

Consider now inferences under the assumption of homogeneous and exogenous θ . First, 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 $3t^2(0.25)^2 = 0.19t^2$. In this case, the endogeneity of θ with respect to t implies that the correct estimate is 300% higher.

Second, suppose that the analyst computes $E[\theta] = 0.5$ by studying responses to high taxes.

⁴³That is, $EB_{neoclassical} = \frac{1}{2}t^2 D(p, t) \frac{\varepsilon_{D,t}}{p+t}$

⁴⁴As shown CLK (and replicated in Proposition 6 for unit demand), the ratio D_t/D_p identifies θ for homogeneous consumers. As shown in the proof of Proposition 7 and implicitly used in the result, it is more generally true that $D_t/D_p = E[\theta|p, t]$ for small t .

Then assuming exogeneity of θ , the analyst would compute the excess burden of tripling the tax to be $3t^2(.5)^2 = 0.75t^2$. In this case, endogeneity of θ implies that the correct excess burden estimate is 20% higher.

6.3 Covariance of Underreaction and Income

Finally, we perform a rough calibration of how the covariance between income and θ affects welfare. To maintain the same benchmark that we have with deadweight loss estimates, we assume that $\lambda = \int g_i = 1$. Under this assumption, the welfare impact of a small tax t is exactly equal to the efficiency loss of a small tax t when the covariance term $Cov[g_i, 2\theta - \theta^2]$ equals zero.

To estimate the covariance, we begin with a noisy estimate $\tilde{\theta}_{ik} = \theta_{ik} + \epsilon_{ik}$ for each person-product pair. Then assuming $E[\epsilon_{ik}] = 0$ and $Cov[g_i, Var[\epsilon_{ik}]] = 0$, it follows that $Cov[g_i, 2\theta - \theta^2] = Cov[g_i, 2\tilde{\theta} - \tilde{\theta}^2]$. We compute the $\tilde{\theta}_{ik}$ as we did in Section 4.6 by first estimating order effects, then partialing those out from y_{ik} , and then dividing by τ_i . Because the θ_{ik} are measured with noise, for this rough calculation we drop all observation of $\tilde{\theta}_{ik}$ that are greater in absolute value than 5, which roughly corresponds to dropping the top and bottom 5% of the $\tilde{\theta}_{ik}$ estimates.

As a final step, we specify the welfare weights g_i . Roughly consistent with Saez (2001), we set $g(Z_i) \propto 1/(10000 + Z_i)$. We add the constant 10000 to the denominator because without the constant consumers with zero (or close) to zero household income receive infinite weight. Under these assumptions, the covariance term is -0.18 when pooling the standard and triple tax arms.

Assuming homogeneous θ and using the estimate $E[\theta] \approx 0.25$ from the standard tax arm, $W(t) - W(0) \approx \frac{1}{2}t^2 E[\theta]^2 D_p \approx 0.03t^2 D_p$. The covariance adds a welfare loss of $\frac{1}{2}t^2(0.18)D_p = 0.09t^2 D_p$, which thus increases the welfare loss by 300%.

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 is greatest amongst low-income consumers, this increases the regressivity of the tax burden and hinders redistributive goals. 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 accommodate any such settings with only minimal modification. This framework can thus serve as

a template for robust behavioral welfare analysis in the presence disaggregated pricing or shrouded attributes. As a concrete example, in future work we aim to directly implement this framework in an experimental study of the impact of heterogeneous inattention to shipping costs.

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 elicitation of demand curves, we worry that this mechanism could trigger a different psychology than simply deciding whether or not to purchase a given item. Within our experimental framework, we plan on conducting future work to establish the robustness of our results to potentially more natural mechanisms, such as multiple price lists or individual purchasing decisions.

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. Furthermore, our decision to exclude participants who fail comprehension checks—done in order to exclude confounds of our measures of underreaction to taxes driven by participants not (carefully) reading the instruction—could drive additional selective processes.⁴⁵ In general, 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. However, we caution that the magnitude of this variation, and particularly its covariation with demographic characteristics, could vary across different populations. 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.

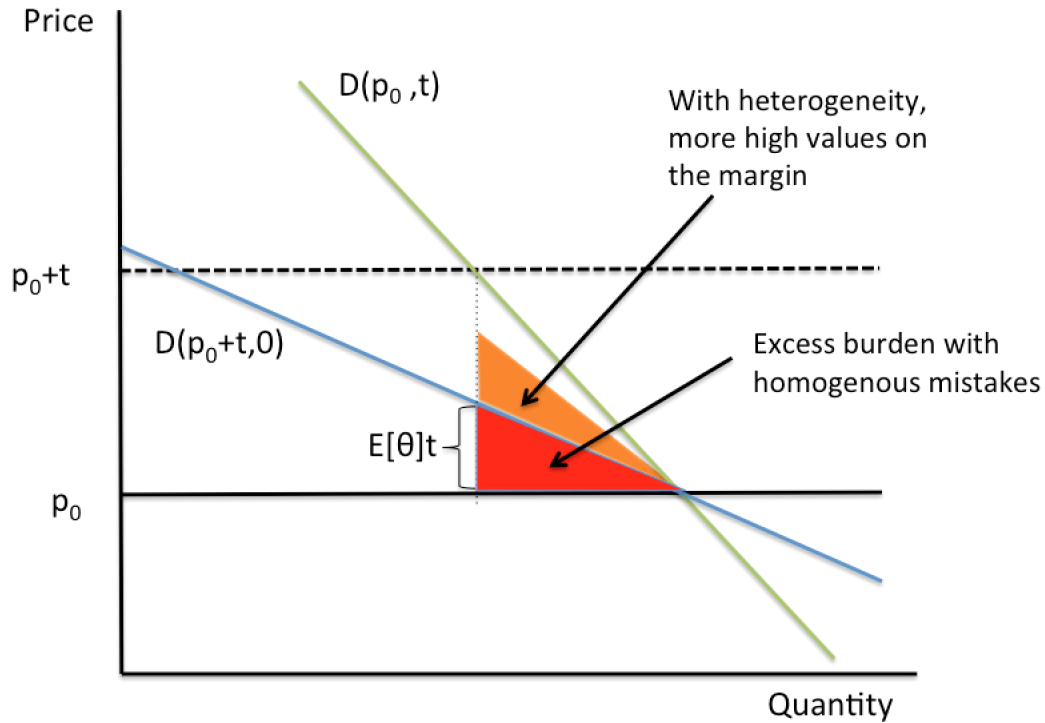
As this agenda progresses, it will both benefit from, and inform, the explicit modeling of the psychology of bounded-rationality. In principal, refined and vetted models of attention would place useful structure on our forecasts of heterogeneity in mistakes, and thus the corresponding implications for welfare. Some of our results—e.g., the debiasing that occurs with increased incentives, and the association between tax numeracy and tax mistakes—are consistent with existing models of bounded rationality (e.g., Gabaix 2014). Some of our results—e.g., that the poor are most susceptible to tax mistakes even after controlling for proxies of tax numeracy and financial literacy—are less consistent with existing rational-attention models, and demonstrate the need for further refinement. In future work, we aim to focus on the economic and psychological determinants of this

⁴⁵However, we do show that our main results are robust to alternative manners of defining our sample for analysis, and to controlling for selection induced by this requirement. See section 4.7.

type of underreaction to shrouded attributes, working towards the ultimate goal of a systematic understanding of the mechanisms in play. This can provide a direct channel linking individual-level modeling of consumer psychology to the measurement of welfare.

Figures

Figure 1: Illustration of efficiency cost computation



Notes: This figure provides a graphical representation of the basic efficiency cost of taxation formula in Proposition 1. If consumers optimize perfectly, excess burden is given by the triangle under the demand curve in green. If consumers underreact to taxes but are homogeneous in their misoptimization, excess burden is given by the red triangle, the height of which is $E[\theta]t$. Heterogeneity creates additional excess burden, represented by the orange triangle, whose height is $tVar[\theta]/E[\theta]$.

Figure 2: Experimental Design

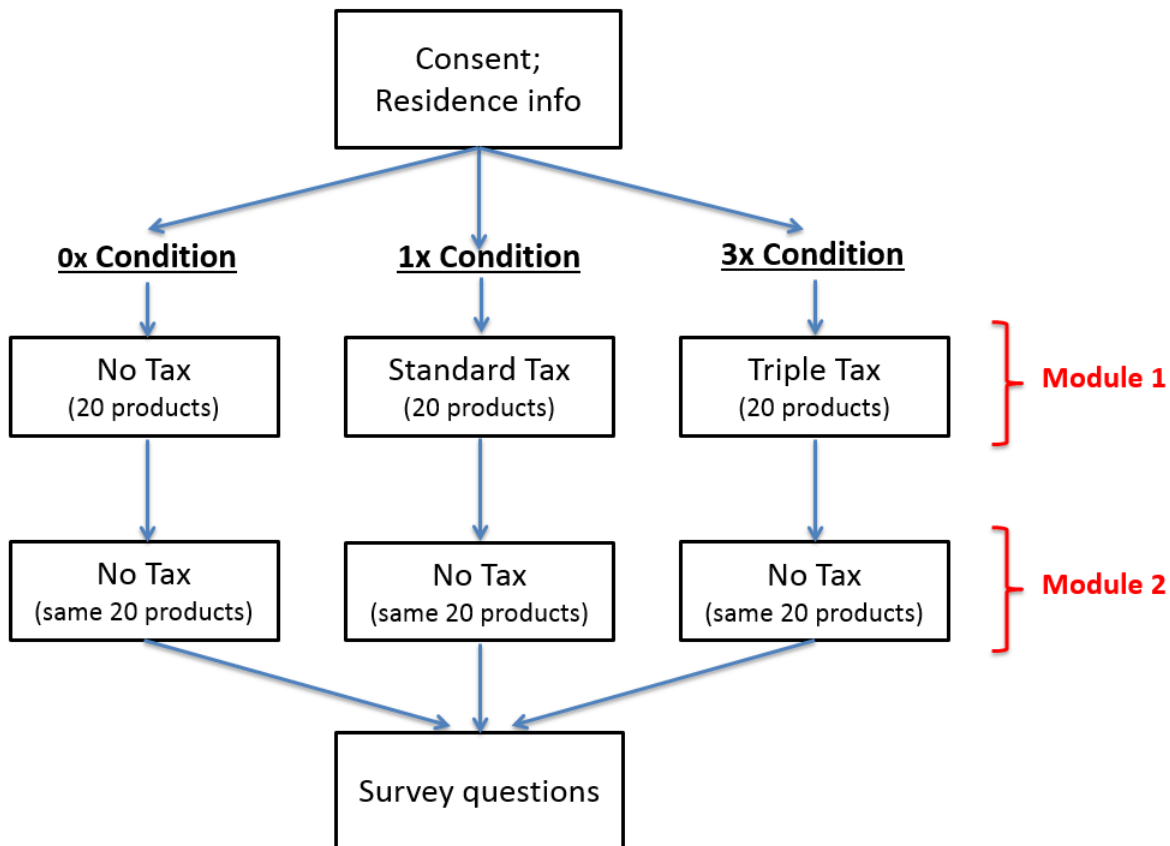


Figure 3: Decision format

(a) Tax Module

Wharton
UNIVERSITY OF PENNSYLVANIA

Carnation Home Fashions Hotel Collection 8-Gauge Vinyl Shower Curtain Liner with Metal Grommets, Monaco Blue



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 repellent 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.

Please enter the highest tag price at which you would buy this product:
(You can drag the slider, or use left and right arrow keys to adjust it by \$0.01 increments)

Tag Price: **\$10.95**




[\[link to instructions page\]](#)

(b) No Tax Module

Wharton
UNIVERSITY OF PENNSYLVANIA

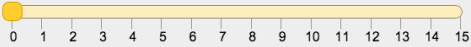
Energizer AA Batteries max Alkaline 20-Pack



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.

Please enter the highest tag price at which you would buy this product if no sales tax were to be applied at the register:
(You can drag the slider, or use left and right arrow keys to adjust it by \$0.01 increments)

Tag Price:

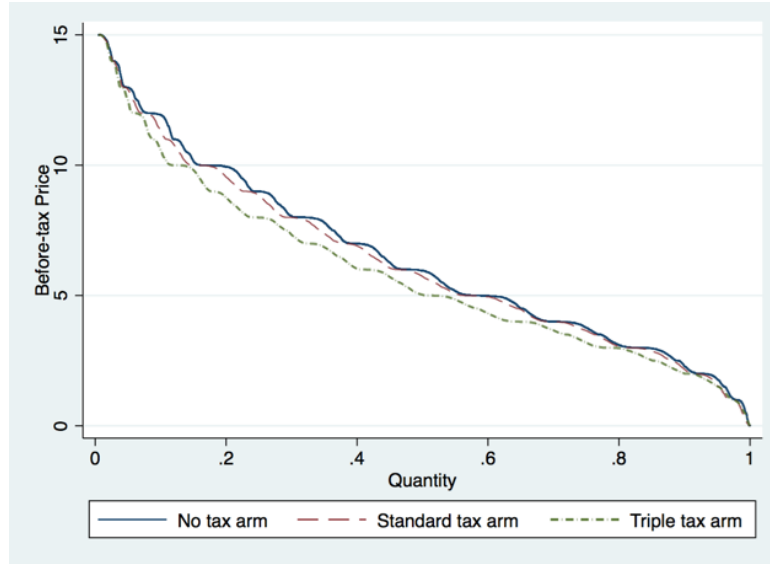


[\[link to instructions page\]](#)

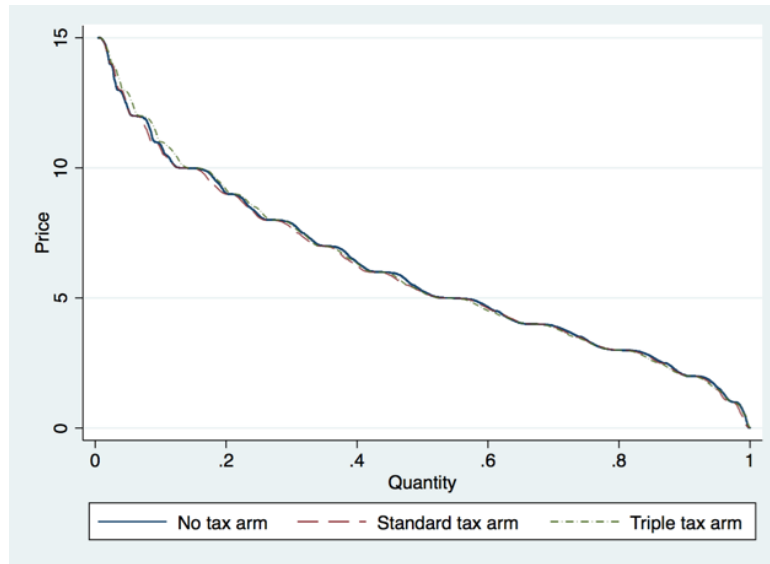
Notes: Panel a shows an example of a pricing decision from modules where taxes apply. Consumers indicate the highest tag price at which they would buy the product. As in typical shopping environments—and as was explained in the experimental instructions—the final price that applies at "check out" is 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 4: 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



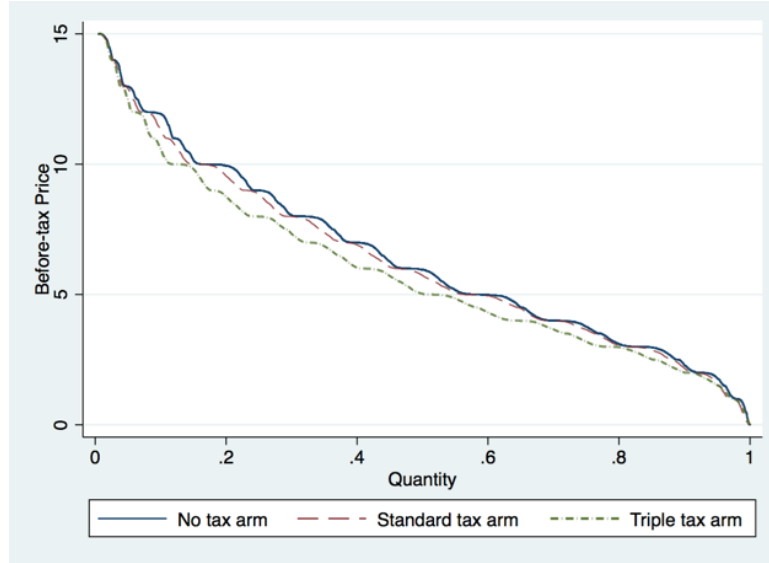
(b) Demand in the second stage, where there are no additional sales taxes in any arm



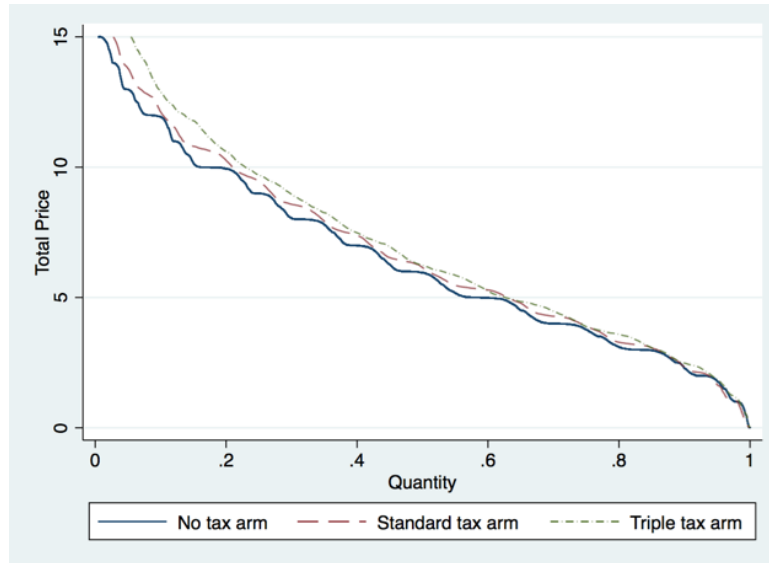
Notes: This figure plots demand curves from the first and second modules of the experiment, averaging across all 20 products. In the first stage, consumers face either no (additional) taxes, standard taxes, or triple their standard taxes. In the second stage, consumers in all three arms face no additional 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 $D_k^{C,m}(p)$ 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 $D_{avg}^{C,m}(p) := \frac{1}{20} \sum_k D_k^{C,m}(p)$.

Figure 5: 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

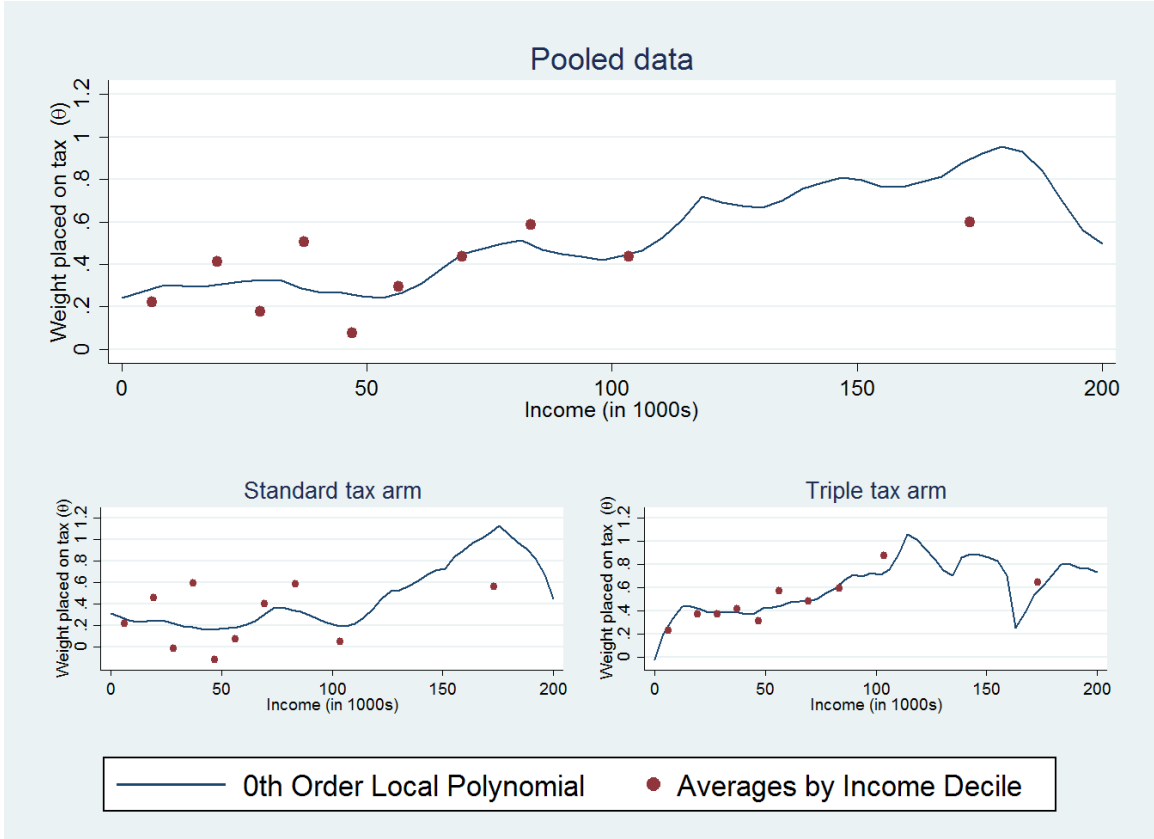


(b) Demand as a function of after-tax price in the first stage of the experiment



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 $D_k^{C,m}(p)$ 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 $D_{avg}^{C,m}(p) := \frac{1}{20} \sum_k D_k^{C,m}(p)$.

Figure 6: Average θ (weight placed on tax) by income, by experimental arm

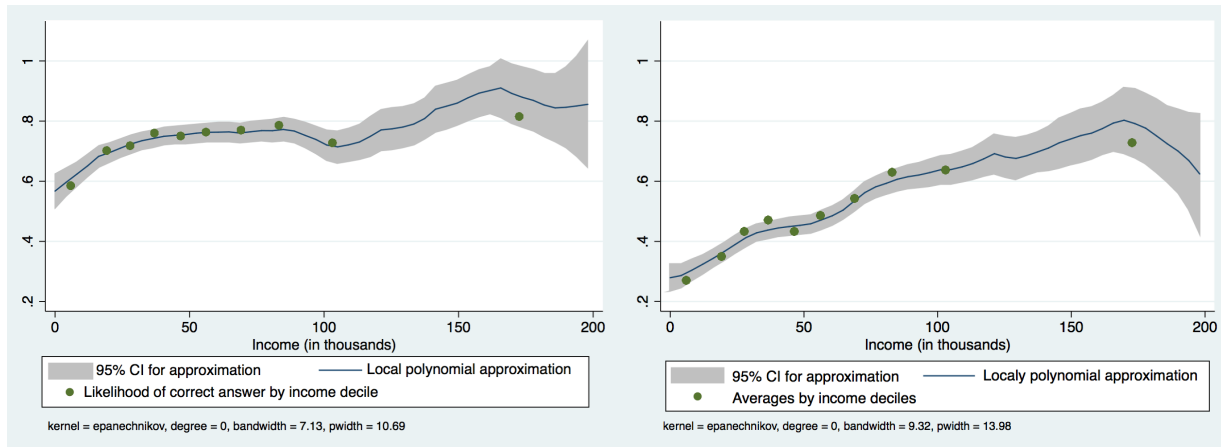


Notes: This figure provides a graphical summary of how θ (weight placed on tax) varies by household income. The top panel presents analysis pooling across treatment arms, while the bottom two panels restrict the data to the standard tax and triple tax arms, respectively. Our measure of θ is calculated for each consumer-product pair, and given by $\frac{\log(p_2^{i^k}) - \log(p_1^{i^k}) - \hat{\beta}\Gamma}{\tau_i}$, where $\hat{\beta}$ is our estimate of order effects identified from the no tax arm. We fit a local polynomial regression (of degree 0) weighted by the Epanechnikov kernel. For each income decile, we also plot our estimate of average θ against the average income in that decile.

Figure 7: Differences in numeracy and financial literacy across income levels

(a) Ability to compute tax by income

(b) Financial literacy by income



Notes: These figures provide a graphical summary of how tax numeracy and financial literacy vary across income levels. Panel a displays the probability that the respondent correctly calculated the tax on an \$8 item (using his *subjective* belief about their sales tax rate). Panel b displays the probability that the respondent correctly answered all of the Lusardi and Mitchell “big three” financial literacy questions (question text available in appendix F). In each figure, we present the average of the dependent variable in each income decile, as well as 0th order local polynomial regressions (using the Epanechnikov kernel).

Tables

Table 1: Demographics by experimental arm

	All data	No tax arm	Standard tax arm	Triple tax arm	p -value of F-test for equality across arms
Age	50.49 (14.63)	50.81 (14.27)	50.41 (14.85)	50.20 (14.82)	$p = 0.648$
Household Income	61.48 (47.79)	60.71 (48.65)	61.72 (47.28)	62.16 (47.32)	$p = 0.668$
Household size	2.40 (1.53)	2.40 (1.60)	2.38 (1.50)	2.42 (1.46)	$p = 0.881$
Male	0.48	0.47	0.51	0.47	$p = 0.193$
Married or domestic partnership	0.61	0.61	0.60	0.61	$p = 0.353$
Education					
Highschool degree or higher	0.96	0.96	0.96	0.96	$p = 0.843$
College degree or higher	0.41	0.41	0.40	0.40	$p = 0.863$
Postgraduate education	0.17	0.17	0.16	0.16	$p = 0.410$
Ethnicity					
Asian	0.03	0.03	0.04	0.03	$p = 0.862$
Caucasian	0.77	0.76	0.77	0.78	$p = 0.549$
Hispanic	0.03	0.04	0.03	0.03	$p = 0.476$
African American	0.07	0.08	0.07	0.08	$p = 0.826$
Other	0.02	0.03	0.02	0.02	$p = 0.393$
Tax rate in city of residence	7.32 (1.15)	7.36 (1.16)	7.31 (1.13)	7.30 (1.15)	$p = 0.376$
N	3000	1103	983	914	

Notes: This table summarizes the demographics in each of the three arms in our final sample for analysis. 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 there are no significant differences in demographics between Arm 1 vs. Arm 2 (F -test $p = 0.38$), Arm 2 vs. Arm 3 (F -test $p = 0.36$) or Arm 1 vs. Arm 3 (F -test $p = 0.44$)

Table 2: Estimates of average θ (weight placed on tax) by condition

	(1)	(2)	(3)
	All	$p^2 \geq 1$	$p^2 \geq 5$
Std. tax avg. θ	0.265** (0.111)	0.252*** (0.095)	0.227** (0.093)
Triple tax avg. θ	0.482*** (0.045)	0.475*** (0.039)	0.535*** (0.041)
Observations	60000	58517	32816
Difference p -val	0.033	0.013	< 0.001

Notes: This table displays method of moments estimates of average θ by experimental arm. 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. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Average θ (weight placed on tax) for different ranges of module 2 price p_2

	Without individual θ fixed effects			With individual θ fixed effects		
	(1) Standard	(2) Triple	(3) Pooled	(4) Standard	(5) Triple	(6) Pooled
High p_2 bin	0.113 (0.185)	0.148** (0.074)	0.155** (0.074)	0.166 (0.152)	0.069 (0.053)	0.072 (0.053)
Middle p_2 bin	-0.098 (0.147)	0.118** (0.054)	0.124** (0.054)	-0.078 (0.101)	0.097** (0.038)	0.105*** (0.038)
Std. tax cons.	0.269* (0.140)		0.156 (0.098)			
Triple tax cons.		0.401*** (0.054)	0.394*** (0.054)			
Observations	40690	39397	58517	40690	39397	58517

Notes: This table displays method of moments estimates of differences in average θ between 1) choices for which the module 2 price p_2 is in the $[5, 10]$ range (“Middle bin”) versus the $[1, 5]$ range, and 2) choices for which the module 2 price p_2 is in the $[10, 15]$ range (“High bin”) versus the $[1, 5]$ range. 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 $\bar{\theta} = \alpha_0^{1x} \mathbf{1}_{1x} + \alpha_0^{3x} \mathbf{1}_{3x} + \alpha_{p_2 \in [5,10]} \mathbf{1}_{p_2 \in [5,10]} + \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}$, where $\bar{\theta} := E \left[\frac{\log(1+\theta\tau)}{\tau} \right]$. 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 θ fixed effects, estimating the model $\theta_{ik} = \theta_i + \alpha_{p_2 \in [5,10]} \mathbf{1}_{p_2 \in [5,10]} + \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}$. We assume that $\alpha_{p_2 \in [5,10]}$ and $\alpha_{p_2 \geq 10}$ do not change across the standard and triple tax arms.

All specifications condition on module 2 price (p_2) being greater than 1, and include a second estimating equation that controls for order effects (see section 4.3 for details). We set up 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. Cluster-robust standard errors (at the subject level) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Average θ (weight placed on tax) by ability to correctly compute tax, financial sophistication, and household income quartiles

	(1)	(2)	(3)	(4)
Compute tax correctly	0.218** (0.087)			0.168* (0.088)
Financially sophisticated		0.243*** (0.078)		0.160** (0.081)
Inc. quartile 2			0.038 (0.109)	0.013 (0.110)
Inc. quartile 3			0.234** (0.113)	0.180 (0.114)
Inc. quartile 4			0.338*** (0.111)	0.272** (0.115)
Std. tax cons.	0.105 (0.115)	0.124 (0.102)	0.091 (0.115)	-0.068 (0.133)
Triple tax cons.	0.318*** (0.074)	0.358*** (0.052)	0.320*** (0.080)	0.161 (0.103)
Observations	58005	58517	58517	58005

Notes: This table shows how average θ varies by ability to compute the tax correctly, financial sophistication, and income quartiles. 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. Data is pooled from both the standard and triple tax conditions.

The model estimated in column (4) is $\bar{\theta} = \alpha_0^{1x} \mathbf{1}_{1x} + \alpha_0^{3x} \mathbf{1}_{3x} + \sum_{j \in \{2,3,4\}} \alpha_{Qj} \mathbf{1}_{Qj} + \alpha_{\text{compute correctly}} \mathbf{1}_{\text{compute correctly}} + \alpha_{\text{sophisticated}} \mathbf{1}_{\text{sophisticated}}$, where $\bar{\theta} := E \left[\frac{\log(1+\theta\tau)}{\tau} \right]$. The models estimates in columns (1)-(3) are analogous but include only the subset of covariates described in those columns.

All specifications condition on module 2 price (p_2) being greater than 1, and include a second estimating equation that controls for order effects (see section 4.3 for details). We set up 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. Cluster-robust standard errors (at the subject level) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Predictiveness of self-classifying survey responses

	(1)	(2)
	Standard	Triple
“Yes” average θ	1.103*** (0.277)	0.936*** (0.103)
“A little” average θ	0.439*** (0.110)	0.622*** (0.048)
“No” average θ	-0.172 (0.138)	0.047 (0.056)
Observations	40690	39397

Notes: This table displays method of moments estimates of average θ by consumers’ responses to the self-classifying survey questions. 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. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on module 2 price (p_2) being greater than 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Lower bound estimates for the expected conditional variance of θ (weight placed on tax)

	Standard	Triple
Lower bound estimate	0.133	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 estimates lower bounds for $E_{p_1, \tau}[Var[\theta|p_1\tau]]$ in both the standard and triple tax arms. 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. The statistic we estimate is (14), via the empirical moments described in Appendix E.4. 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).

Table 7: Summary of illustrative welfare calculations from Section 6

Type of Welfare Calculation	Type of Variation	Increase in welfare costs	
		Standard US sales taxes	Triple taxes
Excess burden from tax	Variance of θ due to individual differences	>200%	>40%
Welfare with redistributive motives, with welfare weights $g(Z) \propto \frac{1}{10000+Z}$, and $\lambda = 1$	Covariance of θ and household income	300%	75%
Excess burden from increasing tax to triple its value	Variation in θ due to changes in tax rate: standard tax baseline.		300%
Excess burden from increasing tax to triple its value	Variation in θ due to changes in tax rate: triple tax baseline.		20%

Notes: This table summarizes the illustrative welfare calculations in Section 6. The benchmark for all of these computation is the estimates that would be obtained under the assumption that θ is homogeneous and does not vary with tax rates. The first row computes efficiency costs of the taxes, under the assumption of quasilinear utility and perfectly elastic supply.

The second row computes the welfare impact of the covariance between θ and household income. Welfare is given by $W = \int g_\omega(Z_\omega + (v - p - t)\mathbf{1}_x)dF + \lambda D$, where λ is the marginal value of public funds and g_ω is the weight on consumers with wealth Z_ω (see Section 2.3 for details). In the second row, the welfare weight corresponding to a consumer with income Z is $g = \frac{1}{10000+Z}$. The marginal value of public funds is set equal to $\lambda = 1$, to increase comparability to the excess burden calculations.

The last row of the table presents calculations for computing the excess burden of increasing standard US sales taxes to triple their current size. We again assume quasilinear utility and perfectly elastic supply. The two different effects presented arise from assuming that underreaction is fixed at the levels seen under standard taxes, or from assuming that underreaction is fixed at the levels seen under triple taxes.

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*, 2014.
- **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.
- Andersen, Steffen, Seda Ertac, Uri Gneezy, Moshe Hoffman, and John A. List**, “Stakes Matter in Ultimatum Games,” *American Economic Review*, 2011, *101*, 3427–3439.
- 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.
- 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.
- Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian**, “Default Stickiness among Low-Income Individuals,” *working paper*, 2012.
- 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.
- Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer**, “Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games,” *Review of Economic Studies*, 2014, *81* (3), 944–970.
- Brown, Jeffrey R., Arie Kapteyn, Erzo F.P. Luttmer, and Olivia S. Mitchell**, “Cognitive Constraints on Valuing Annuities,” *NBER Working Paper No. 19168*, 2013.
- Bushong, Benjamin, Joshua Schwartzstein, and Matthew Rabin**, “A Model of Relative Thinking,” *Working Paper*, 2015.
- Campbell, John Y.**, “Household Finance,” *The Journal of Finance*, 2006, *61* (4), 1553–1604.
- Caplin, Andrew and Mark Dean**, “The Behavioral Implications 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.
- Charness, Gary, Edi Karni, and Dan Levin**, “On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda,” *Games and Economic Behavior*, 2010, *68*, 551–556.

- Chetty, Raj**, “The Simple Economics of Salience and Taxation,” Technical Report, National Bureau of Economic Research 2009.
- , “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods,” *Annual Review of Economics*, 2009, 1, 451–488.
- , “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.
- , —, **Soren Leth-Petersen**, **Torben Nielsen**, and **Tore Olsen**, “Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark,” *Quarterly Journal of Economics*, 2014, 139 (3), 1141–1219.
- 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.
- and **Joshua M. Pollet**, “Investor Inattention and Friday Earnings Announcements,” *Journal of Finance*, 2009, 64, 709–749.
- Farhi, Emmanuel** and **Xavier Gabaix**, “Optimal Taxation with Behavioral Agents,” Working Paper 21524, National Bureau of Economic Research September 2015.
- Feldman, Naomi E.**, **Peter Katuscak**, and **Laura Kawano**, “Taxpayer Confusion over Predictable Tax Liability Changes: Evidence from the Child Tax Credit,” *Finance and Economics Discussion Series Working Paper*, 2013.
- 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).
- Gerardi, Kristopher**, **Lorenz Goette**, and **Stephan Meier**, “Numerical Ability Predicts Mortgage Default,” *Proceedings of the National Academy of Sciences*, 2013, 110 (28), 11267–1271.
- Goldin, Jacob**, “Optimal Tax Salience,” *Journal of Public Economics*, Forthcoming.
- Handel, Benjamin R** and **Jonathan T Kolstad**, “Health Insurance for “Humans”: Information Frictions, Plan Choice, and Consumer Welfare,” *American Economic Review*, 2015, 105 (8), 2449–2500.
- 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.
- Hendren, Nathaniel**, “Private Information and Insurance Rejections,” *Econometrica*, 2013, *81* (5), 1713–1762.
- 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.
- Reck, Daniel H.**, “Taxes and Mistakes: What’s in a Sufficient Statistic?,” *Working Paper*, 2014.
- Rees-Jones, Alex**, “Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence from the Residency Match,” *Working Paper*, 2015.
- **and Dmitry Taubinsky**, “Heuristic Perceptions of the Income Tax: Evidence and Implications,” *Working Paper*, 2015.
- 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.
- van Rooij, Maarten, Annamria Lusardi, and Rob Alessie**, “Financial Literacy, Retirement Planning, and Household Wealth,” *Economic Journal*, 2012, *122* (560), 449–478.
- 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.

Appendices (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 6. 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 6 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 3. 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 7. 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 ,

$$\underline{d} = \frac{1}{2} \left(\frac{D_t(p, \Delta t)}{D_p(p, \Delta t)} \right)^2 (\Delta t)^2 D_p \quad (16)$$

$$\bar{d} = \frac{1}{2} \left(\frac{D_t(p, \Delta t)}{D_p(p, \Delta t)} \right) \bar{\theta} (\Delta t)^2 D_p \quad (17)$$

Proposition 7 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 (16), and it is largest when all consumers either have $\theta = \bar{\theta}$ or $\theta = 0$, which corresponds to \bar{d} in (17).

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 large income effects in a discrete choice model is a somewhat awkward and not a standard analysis even the standard model of full tax salience. We consider with utility functions $U_\omega(x, y) = u_\omega(x) + v_\omega(y)$, where $\omega \in \Omega$. Individuals choose x to maximize $U(x, Z - (p + \theta t)x)$. Our definition of θ here is analogous to the definition of θ for the binary demand model: it is the amount by which a person underreacts to the tax. Again, θ has an immediate choice-based interpretation. If $x_i^*(p, t)$ is an individual's choice of x given a tax t and a price p , and if $q^*(t)$ is the value of p for which $x_i^*(q^*, 0) = x_i^*(p, t)$, then $\theta_i(p, t) = (q^*(t) - p)/t$. It is straightforward to see that $\theta_i = \left(\frac{dx_i^*}{dt}\right) / \left(\frac{dx_i^*}{dp}\right)$. Analogously, we can define, as in CLK, a ‘‘compensated θ ’’ as a ratio of compensated demand elasticities: $\theta_i^c = \left(\frac{dx_i^*}{dt} + x_i^* \frac{dx_i^*}{dZ}\right) / \left(\frac{dx_i^*}{dp} + x_i^* \frac{dx_i^*}{dZ}\right)$. CLK show that

$$EB_i(t) \approx -\frac{1}{2}t^2\theta_i^c \left(\frac{dx_i^*}{dt} + x \frac{dx_i^*}{dZ}\right)$$

Assuming for simplicity that θ_i^c is not correlated with the compensated demand response $\frac{dx_i^*}{dp} + \frac{dx_i^*}{dZ}$ (which occurs, for example, if the true utility function is the same for all consumers), we have that

$$EB(t) \approx -\frac{1}{2}t^2 \left(E[\theta_i^c|p, t]^2 + Var[\theta_i^c|p, t]\right) \varepsilon_p^c \frac{D}{p+t} \quad (18)$$

More generally, the formula in (18) will include an additional term given by $-\frac{1}{2}t^2 Cov\left[\theta_i^c, \frac{dx_i^*}{dp} + x \frac{dx_i^*}{dZ}\right]$.

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 $j = 1, \dots, J$ such that type j consumers reacts to a tax t as if it was $\theta(t)t$. Second, we assume that γ is independently distributed of v and θ .

The policymaker's objective function is to maximize

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

We now characterize optimal taxes in this more general model.

Proposition 8. *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*

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

The general formula in part 1 of Proposition 8, 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 simply trades off the deadweight loss computed in Proposition 1 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 1 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 from 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*

$$t = \frac{\lambda - 1 + \bar{\gamma}\varepsilon_{D,t}}{(\lambda - 1)(\varepsilon_{D,t} - 1) + E[\theta|t]\varepsilon_{D,t}} \quad (19)$$

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

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

In this last special case, equation (20) 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, \dots, 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, \dots, 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 (20) 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 9. *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, $\frac{d}{d\rho}W(t^*) = \frac{2\lambda - 1}{2}m(t^*)^2 - \bar{\gamma}m t^*$, 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

$$\bar{\gamma} > \frac{(2\lambda - 1)(1 - \lambda)}{\rho m}. \quad (21)$$

Equation (21) 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 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 Γ denote the vector of order dummies and X the vector of covariates of θ .

Proposition 10. 1. Assume that for both the standard and triple tax arms ($C=1x$ or $C=3x$), $E[y_{ik}|\pi = 1, C, \Gamma, X] = E[y_{ik}|\pi = 1, C, \Gamma]$. Then for the standard and triple tax arms ($C=1x$ or $C=3x$),

$$E[y_{ik}|\pi = 1, C, \Gamma, X] = E[y_{ik}|\pi = 1, C, \Gamma, \theta_{ik} = 0] + E[\tilde{\theta}|\pi = 1, C, \Gamma, X]$$

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

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

3. Assume that $Pr(\pi = 1|\eta, C=3x) \leq Pr(\pi = 1|\eta, C=1x) \forall \eta$. Then

$$\begin{aligned} & \int E[\tilde{\theta}|\eta, \pi = 1, C=3x, X]dF(\eta|\pi = 1, C=1x, X) - \int E[\tilde{\theta}|\eta, \pi = 1, C=1x, X]dF(\eta|\pi = 1, C=1x, X) \\ & \geq \frac{Pr(\pi = 1|C=3x)}{Pr(\pi = 1|C=1x)} E[\tilde{\theta}|\pi = 1, C=3x, X] - E[\tilde{\theta}|\pi = 1, C=1x, X] \end{aligned} \quad (22)$$

Part 1 of the proposition shows 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 a slight extension of the approach taken by Jones and Mahajan (2015) and Behaghel et al. (2009) to bound treatment effects in experiments with attrition—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

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 $E[\theta|\pi = 1, C=3x, X]$ must be deflated by the ratio $\frac{Pr(\pi=1|C=3x)}{Pr(\pi=1|C=1x)}$ 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 (22), 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 $E[\theta|\pi = 1, C=3x]$ and $E[\theta|\pi = 1, C=1x]$. The third moment condition employs the full sample to estimate $\frac{Pr(\pi=1|C=3x)}{Pr(\pi=1|C=1x)}$. We use these estimate to derive the lower-bound (22), and we use the delta method to obtain standard errors.

D Proofs of Propositions

We introduce one new piece of notation for the proofs. We let $\tilde{EB}(t, \tilde{F})$ denote the excess burden of a tax t , given a distribution $F(\theta, v)$ that does not depend on the tax. Thus \tilde{EB} is defined as a function of the tax t and an exogenous distribution of \tilde{F} .

Proof of Proposition 1 Follows from Proposition 2.

Proof of Proposition 2 Set $\tilde{F}(\theta, v) = F(\theta, v|t)$, and note that by definition, $EB(t) = \tilde{EB}(t, \tilde{F})$. Let p_0 be the initial price and let p_1 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

$$\begin{aligned}
 \tilde{EB}(t, \tilde{F}) &= \overbrace{\left[\int_{v \geq p_0} (v - p_0) d\tilde{F} - \int_{v \geq p_1 + \theta t} (v - p_1 - t) d\tilde{F} \right]}^{\text{Equivalent variation in wealth for consumers}} - \overbrace{\int_{v \geq p_1 + \theta t} t d\tilde{F}}^{\text{Change in government revenue}} \\
 &+ \overbrace{(p_0 x_0^* - C(x_0^*)) - (p_1 x_1^* - C(x_1^*))}^{\text{Change in producer profits}} \\
 &= \int_{v \geq p_0} (v - p_0) d\tilde{F} - \int_{v \geq p_1 + \theta t} (v - p_1) d\tilde{F} + (p_0 x_0^* - C(x_0^*)) - (p_1 x_1^* - C(x_1^*))
 \end{aligned}$$

Now by the multidimensional Leibniz rule,

$$\begin{aligned}
\frac{d}{dt} \int_{v \geq p_1 + \theta t} (v - p_1 - t) d\tilde{F} &= - \int \left[(\theta t) \frac{d}{dt} (\theta t + p_1) \right] d\tilde{F}(\theta, v|v = p_1 + \theta t) \\
&+ \int_{v \geq p_1 + \theta t} \left(-\frac{dp_1}{dt} \right) d\tilde{F} \\
&= - \int (\theta t) \left(\theta + \frac{dp_1}{dt} \right) d\tilde{F}(\theta, v|v = p_1 + \theta t) \\
&+ \int_{v \geq p_1 + \theta t} \left(-\frac{dp_1}{dt} \right) d\tilde{F} \\
&= - tE_{\tilde{F}} \left[\theta^2 + \theta \frac{dp_1}{dt} |v = p_1 + \theta t \right] \int d\tilde{F}(\theta, v|v = p_1 + \theta t) - x_1^* \frac{dp_1}{dt} \\
&= -tE_{\tilde{F}}[\theta]E \left[\theta + \frac{dp_1}{dt} |v = p_1 + \theta t \right] \int d\tilde{F}(\theta, v|v = p_1 + \theta t) \\
&- tVar_{\tilde{F}}[\theta|v = p_1 + \theta t] \int d\tilde{F}(\theta, v|v = p_1 + \theta t) - x_1^* \frac{dp_1}{dt} \\
&= tE_{\tilde{F}}[\theta|p_1, t] \frac{d}{dt} x_1^* + tVar_{\tilde{F}}[\theta|p_1, t] D_p(p_1, t) - x_1^* \frac{dp_1}{dt} \tag{23}
\end{aligned}$$

To arrive to the final equation in (23) from the preceding equation, we use the fact that

$$\begin{aligned}
\frac{d}{dt} x_1^* &= \frac{d}{dt} \int_{v \geq p_1 + \theta t} d\tilde{F} \\
&= - \int \frac{d}{dt} (p_1 + \theta t) d\tilde{F}(\theta, v|v = p_1 + \theta t) \\
&= - \int \left(\theta + \frac{d}{dt} p_1 \right) d\tilde{F}(\theta, v|v = p_1 + \theta t)
\end{aligned}$$

Next, the Envelope Theorem implies that

$$\frac{d}{dt} (p_1 x_1^* - C(x_1^*)) = x_1 \frac{dp_1}{dt}$$

Putting this together, we thus have that

$$\frac{d}{dt} \tilde{E}B(t, \tilde{F}) = tE_{\tilde{F}}[\theta|v = p_1 + \theta t] \frac{d}{dt} x_1^* + tVar_{\tilde{F}}[\theta|v = p_1 + \theta t] D_p(p_1, t) \tag{24}$$

Assuming that $E[\theta|p_1, t]$, $Var[\theta|p_1, t]$, D , and x_1^* are smooth, it follows that

$$\frac{d^2}{dt^2} \tilde{E}B(t, \tilde{F}) = E_{\tilde{F}}[\theta|v = p_1 + \theta t] \frac{d}{dt} x_1^* + Var_{\tilde{F}}[\theta|v = p_1 + \theta t] D_p(p_1, t) + O(t) \tag{25}$$

where $O(t)$ represents all terms of order t or higher (as $t \rightarrow 0$). A Taylor expansion thus implies that

$$EB(t) = \tilde{E}B(t, \tilde{F}) - \tilde{E}B(0, \tilde{F}) = -\frac{1}{2}t^2 \left[E_{\tilde{F}}[\theta|v = p_1 + \theta t] \frac{d}{dt} x_1^* + Var_{\tilde{F}}[\theta|v = p_1 + \theta t] D_p(p_1, t) \right] + O(t^3) \tag{26}$$

where $O(t^3)$ represents all terms of t^3 or higher (as $t \rightarrow 0$).

Proof of Proposition 3 Let $\tilde{F}_2(v, \theta) = F(\theta, v|t + \Delta t)$ and let $\tilde{F}_1(v, \theta) = F(\theta, v|t)$. Now

$$EB(t + \Delta t) - EB(t) = \tilde{E}B(t + \Delta t, \tilde{F}_2) - \tilde{E}B(t, \tilde{F}_2) + (\tilde{E}B(t, \tilde{F}_2) - \tilde{E}B(t, \tilde{F}_1)). \quad (27)$$

By (24) and (25), it thus follows that

$$\begin{aligned} \tilde{E}B(t + \Delta t, \tilde{F}_2) - \tilde{E}B(t, \tilde{F}_2) &= -t\Delta t [E_{\tilde{F}_2}[\theta|v = p + \theta(t + \Delta t)]D_p + Var_{\tilde{F}_2}[\theta|v = p + \theta(t + \Delta t)]D_p] \\ &\quad - \frac{(\Delta t)^2}{2} [E_{\tilde{F}_2}[\theta^2|v = p + \theta(t + \Delta t)]D_p + Var_{\tilde{F}_2}[\theta|v = p + \theta(t + \Delta t)]D_p] + O(t)O(\Delta t^3) \end{aligned} \quad (28)$$

By (26), it follows that

$$\begin{aligned} \tilde{E}B(t, \tilde{F}_2) - \tilde{E}B(t, \tilde{F}_1) &= -\frac{1}{2}t^2 (E_{\tilde{F}_2}[\theta^2|v = p + \theta t] - E_{\tilde{F}_1}[\theta^2|v = p + \theta t]) + O(t^3) \\ &= -\frac{1}{2}t^2 (E_{\tilde{F}_2}[\theta^2|v = p + \theta(t + \Delta t)] - E_{\tilde{F}_1}[\theta^2|v = p + \theta t]) + O(t^3) + O(\Delta t)O(t^2) \end{aligned} \quad (29)$$

where the last line follows from the approximation $-\frac{1}{2}tE_{\tilde{F}_2}^2([\theta^2|v = p + \theta(t + \Delta t)] - E_{\tilde{F}_2}[\theta^2|v = p + \theta t]) = O(\Delta t)O(t^2)$. Inserting the expressions (28) and (29) into (27), and ignoring terms of order $O(t^3)$, $O(t)O(\Delta t^2)$, $O(\Delta t)O(t^2)$ (by the assumption that t is small like Δt), yields the result.

Proof of Proposition 4

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

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

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

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

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

Proof of Proposition 5 *Step 1.* We first show that $Var[\theta|p, \tau] \geq Var[\phi|p, \tau]$, where $\phi = \frac{\log(1+\theta\tau)}{\tau}$. To that end, note that $\frac{d}{d\theta}\phi = \frac{1}{(1+\theta\tau)} \leq 1$ for $\theta \geq 0$. Thus for any $\theta, \theta', |\theta - \theta'| \geq |\phi(\theta) - \phi(\theta')|$. Thus for each θ'

$$\begin{aligned}
\left(\theta' - \int \theta dF(\theta|p, t)\right)^2 &= \left(\int (\theta' - \theta) dF(\theta|p, t)\right)^2 \\
&\geq \left(\int (\phi(\theta') - \phi(\theta)) dF(\theta|p, t)\right)^2 \\
&\geq \left(\phi(\theta') - \int \phi(\theta) dF(\theta|p, t)\right)^2
\end{aligned}$$

This immediately implies that $E[\text{Var}[\theta|p, \tau]] \geq E[\text{Var}[\phi(\theta)|p, \tau]]$.

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

Step 3. Suppose, first, that the distribution G of (p, τ) is differentiable, with a density function g . Let $\mu(p, \tau) = E[\phi(\theta)|p, \tau]$. Then

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

The Cauchy-Schwarz inequality implies that

$$\begin{aligned} &\int (g(p, \tau)^{1/2} \text{Pr}(R = r|p, \tau)^{1/2})^2 \int (g(p, \tau)^{1/2} \text{Pr}(R = r|p, \tau)^{1/2})^2 (\bar{\theta} - \mu)^2 \\ &\geq \left[\int (\bar{\theta}(r, p, \tau) - \mu(p, \tau)) g(p, \tau) \text{Pr}(R = r|p, \tau) \right]^2 \\ &= \text{Pr}(R = r)^2 (E[\phi(\theta)|R = r] - E[\mu|R = r])^2 \end{aligned}$$

This implies

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

and thus

$$E[\text{Var}[\bar{\theta}|p, \tau]] \geq \sum_r \text{Pr}(R = r)^2 (E[\phi(\theta)|R = r] - E[\mu|R = r])^2. \quad (30)$$

If the distribution of (p, τ) instead takes on only finitely many values, the bound (30) is established using identical reasoning. This then also establishes the bound (30) for any distribution that is a mixture of a differentiable and discrete distribution.

Proof of Proposition 6 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 $EB'(t) = (1 - \theta)tD_t =$

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

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 7 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, $p + \theta(\nu(p, t), t)t = \nu(p, t) = p + \nu(p, t)t$, which implicitly defines the $\theta(\nu, t)$ 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 $D(p, 0) \geq D(p, t) \geq D(p + \bar{\theta}t, 0)$ 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 $\Delta t \rightarrow 0$, note that $EB(\Delta t) \rightarrow -\frac{1}{2}t^2(E[\theta|p, 0]^2 + Var[\theta|p, 0])D_p$ as $\Delta t \rightarrow 0$. 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 $\theta = 0$.

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)	(2)	(3)	(4)
	OLS	0.25 Quantile	0.5 Quantile	0.75 Quantile
1x Arm	-0.09 (0.10)	-0.04 (0.12)	-0.02 (0.08)	-0.08 (0.09)
3x Arm	-0.03 (0.10)	-0.05 (0.12)	-0.03 (0.08)	0.22 (0.16)
Adj. R^2	60000	60000	60000	60000

* 0.1 ** 0.05 *** 0.01

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. Cluster-robust standard errors (at the subject level) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.2 Average θ by State Tax Rate

The main implication of the table below is that when we use only the naturally occurring variation in state tax rates—as in columns (3) and (4) of the table—we do not have enough power to detect how $\bar{\theta}$ changes with the size of the tax. Column (1) is essentially a replication of the endogenous attention result from Table 2, with the only difference being that we present the result as an estimate of how $\bar{\theta}$ changes linearly with the tax rate. The difference between column (1) and column (2) is that while column (1) uses only the experimental arms as an instrument for differences in state tax rates, column (2) uses full variation in the state tax rates. As can be seen, our results are robust across columns (1) and (2). It is only when we move to columns (3) and (4), where we have significantly less variation in tax rates that we lose statistical power.

Table A2: Average θ by state tax rate

Model: $\bar{\theta} = \alpha_0 + \alpha_1\tau$

	(1)	(2)	(3)	(4)
	All data	All data	Standard	Triple
α_0	0.14 (0.14)	0.18 (0.11)	0.33 (0.67)	0.37 (0.26)
α_1	1.53** (0.61)	1.31*** (0.49)	-1.07 (8.69)	0.46 (1.12)
Observations	58517	58517	40690	39397

Notes: Column (1) uses only the experimental variation between the standard- and triple-tax arms as an instrument for tax size. Columns (2)-(4) use all available variation in taxes. Columns (1) and (2) use data from all conditions, Column (3) produces estimates only for the standard tax arm (using data from no-tax and standard-tax arms), Column (4) produces estimates for the triple tax arm (using data from no-tax and standard-tax arms). All specifications condition on $p_2 \geq 1$. All specifications include order-effect dummies. Cluster-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.3 Within-Subject Estimation of Endogenous Attention

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

$$E[\tilde{\theta}_{ik} \mid \mathbf{1}_{p_2 \in [5, 10)}, \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}] = \alpha_i + \alpha_{p_2 \in [5, 10)} X_{ik}^1 + \alpha_{p_2 \geq 10} X_{ik}^2$$

and

$$E[y_{ik}] = \beta \Gamma_{ik} + b_1 X_{ik}^1 + b_2 X_{ik}^2.$$

We set $\bar{\theta}_i = \frac{1}{20} \sum_k \tilde{\theta}_{ik}$ and similarly define $\bar{y}_i = \frac{1}{20} \sum_k y_{ik}$. From this, it follows that

$$E\left[\frac{y_{ik} - \bar{y}_i}{1 + \tau_i \mathbf{1}_{tax}} \mid X, \Gamma\right] = \frac{\beta \Gamma_{ik} - \bar{\beta} \Gamma_i}{1 + \tau_i \mathbf{1}_{tax}} + \alpha_{p_2 \in [5, 10)} (X_{ik}^1 - \bar{X}_i^1) + \alpha_{p_2 \geq 10} (X_{ik}^2 - \bar{X}_i^2) \quad (31)$$

To estimate the parameters, we proceed as before with method of moments, replacing the theoretical moment in (31) with the empirical moment. Note that equation (31) 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.

E.4 Further Details for the Lower-Bound Estimation

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 $Pr(\widehat{R} = r) := \frac{1}{N} \sum \mathbf{1}_{R_i = r}$, 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

$$\widehat{\theta}_r = \frac{1}{N_r} \sum_{i,k} \left[\frac{y_{ik} - \hat{\beta} \Gamma_{ik}}{\tau_i} \mid R_i = r \right].$$

where N_r is the number of consumer-product pairs associated with $R_i = r$, and $\hat{\beta}$ is identified from the no tax arm. Concretely, $\hat{\beta}$ is constructed as the OLS estimate from $E[y_{ik} \mid \Gamma_{ik}] = \beta \Gamma_{ik}$ in the no-tax arm of the experiment.

We estimate $E[y_{ik} \mid C = 0x, \mathbf{p} \times \boldsymbol{\tau}]$, the average order effect in the no tax arm, by

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

where $|\mathbf{p} \times \boldsymbol{\tau}|_{no \text{ tax}}$ 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

$$\widehat{\tilde{\mu}}(\mathbf{p}_1, \boldsymbol{\tau}) := \frac{1}{|\mathbf{p}(\mathbf{p}_1) \times \boldsymbol{\tau}(\boldsymbol{\tau})|} \sum_{(p_1^i, \tau_i) \in \mathbf{p} \times \boldsymbol{\tau}} \frac{y_{ik} - m(\mathbf{p} \times \boldsymbol{\tau})}{\tau_i} \quad (32)$$

Clearly, $\widehat{\tilde{\mu}}(\mathbf{p}_1, \boldsymbol{\tau})$ is an unbiased estimate of $E \left[\frac{y_{ik} - E[y_{ik} \mid C = 0x, \mathbf{p}(\mathbf{p}_1), \boldsymbol{\tau}(\boldsymbol{\tau})]}{\tau_i} \mid \mathbf{p}(\mathbf{p}_1), \boldsymbol{\tau}(\boldsymbol{\tau}) \right]$. 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,

$$E[y_{ik} \mid \theta_{ik}, R_j, \mathbf{p}, \boldsymbol{\tau}] = E[y_{ik} \mid C = 0x, \mathbf{p}, \boldsymbol{\tau}] + E[\log(1 + \theta_{ik} \tau_i) \mid \mathbf{p}, \boldsymbol{\tau}]$$

from which the conclusion follows by rearrangement.

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

$$\sum_{r \in \{L, M, H\}} Pr(\widehat{R} = r) \left(\widehat{\theta}_r - E[\widehat{\mu}_{ik} \mid R = r] \right)^2 \quad (33)$$

By construction, our estimates of $Pr(R = r)$, $\bar{\theta}_r$, and $\tilde{\mu}$ are all unbiased. Note, however, that (33) 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.5 Robustness of Individual Differences Results to the Inclusion of other Demographics

We now replicate Table 4, controlling for other demographic covariates. The demographic variables we include are age, gender, marital status, education (college degree or higher), household size, and ethnicity. We calculate effects for both the triple tax arm and the pooled conditions. The only demographic that appears to relate robustly to θ is age, with older consumers significantly less likely to pay attention to taxes than younger consumers.

Table A3: Average θ by income, ability to compute taxes correctly, and financial sophistication, controlling for demographics

	(1)	(2)	(3)	(4)
Correct	0.263*** (0.092)			0.239** (0.093)
Sophisticated		0.255*** (0.089)		0.199** (0.090)
Quartile 2			0.039 (0.114)	0.015 (0.114)
Quartile 3			0.193 (0.120)	0.137 (0.120)
Quartile 4			0.294** (0.133)	0.231* (0.133)
Age	-0.008** (0.003)	-0.008** (0.003)	-0.007** (0.003)	-0.009*** (0.003)
Male	0.005 (0.086)	-0.042 (0.086)	-0.010 (0.084)	-0.050 (0.085)
Household size	0.025 (0.030)	0.024 (0.029)	0.005 (0.030)	0.014 (0.031)
Married	-0.187** (0.093)	-0.191** (0.092)	-0.124 (0.097)	-0.139 (0.097)
College or higher	0.100 (0.085)	0.056 (0.091)	0.068 (0.091)	0.002 (0.095)
Asian	-0.046 (0.247)	-0.077 (0.252)	-0.099 (0.244)	-0.060 (0.249)
Caucasian	-0.007 (0.146)	-0.003 (0.144)	-0.007 (0.146)	0.008 (0.143)
Hispanic	-0.259 (0.262)	-0.232 (0.259)	-0.315 (0.255)	-0.219 (0.253)
African American	-0.056 (0.193)	-0.017 (0.191)	-0.031 (0.193)	0.033 (0.190)
Observations	53181	53615	53615	53181

Notes: This table replicates Table 4, controlling for other demographic covariates.

E.6 Distribution of Self-Classifying Survey Question Responses

Table A4: 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: This table summarizes the responses of participants in the standard and triple tax arm to the question of whether they would buy at higher tag prices if there was no tax in the first module. The answers are “Yes” ($R = H$), “Maybe a little” ($R = M$), or “No” ($R = L$). The first column summarizes responses in the standard tax arm and the second column summarizes responses in the triple tax arm.

E.7 Individual Differences Results by Experimental Arm

Table A5: Average θ by income, ability to compute taxes correctly, and financial sophistication, by experimental arm

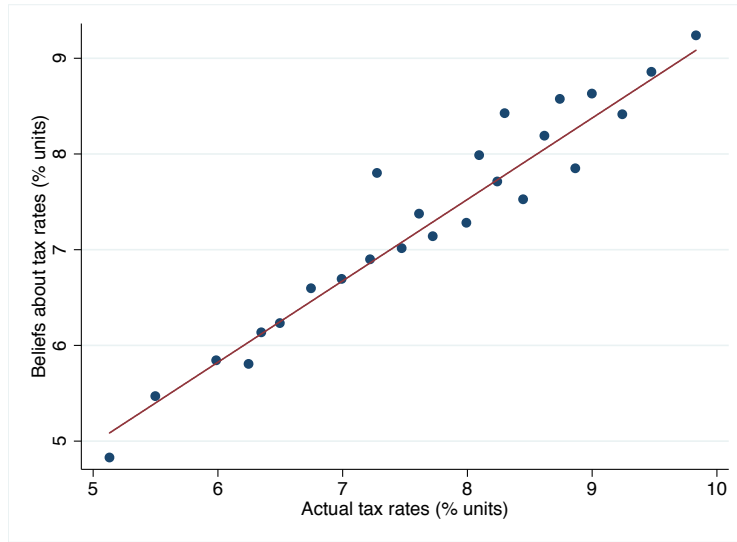
	(1) Triple	(2) Standard	(3) Triple	(4) Standard	(5) Triple	(6) Standard	(7) Triple	(8) Standard
Compute tax correctly	0.219** (0.088)	0.181 (0.216)					0.170* (0.088)	0.139 (0.216)
Financially sophisticated			0.240*** (0.078)	0.321* (0.190)			0.159* (0.081)	0.302 (0.199)
Inc. quartile 2					0.042 (0.109)	0.013 (0.259)	0.020 (0.110)	-0.037 (0.262)
Inc. quartile 3					0.238** (0.113)	0.282 (0.278)	0.181 (0.114)	0.208 (0.280)
Inc. quartile 4					0.350*** (0.112)	0.133 (0.252)	0.281** (0.116)	0.016 (0.264)
Std. tax cons.		0.132 (0.184)		0.088 (0.130)		0.144 (0.177)		-0.038 (0.238)
Trile tax cons.	0.317*** (0.074)		0.358*** (0.052)		0.317*** (0.080)		0.156 (0.103)	
Observations	39121	40255	39397	40690	39397	40690	39121	40255

Notes: This table replicates Table 4, but breaking down the results by experimental arm.

E.8 Graphical Summary of Beliefs About Tax Rates

Figure 8 provides a graphical summary of how beliefs about tax rates vary with actual tax rates. The figure is constructed by partitioning the actual tax rates faced by our consumers into 25 quantiles, and plotting the mean perceived belief corresponding to each quantile against the average of actual tax rates in that quantile.

Figure 8: Perceived vs. actual tax sales tax rates



Notes: This figure plots perceived vs. actual sales tax rates. To construct the figure, we first divide the actual tax rates of our 3000 consumers into 25 quantiles. We then plot the average belief vs. the average actual tax rate for each of the quantiles.

E.9 Replication of Main Results Without Excluding study participants Failing Comprehension Questions

Table A6: Replication of Table 2: method of moments estimates of average θ by condition

	(1)	(2)	(3)
	All	$p_2 \geq 1$	$p_2 \geq 5$
Std. tax avg. θ	0.073 (0.108)	0.114 (0.086)	0.147* (0.085)
Triple tax avg. θ	0.279*** (0.040)	0.294*** (0.032)	0.376*** (0.033)
Observations	84480	82028	44918

Notes: Column (1) uses all data, Column (2) conditions on $p_2 \geq 1$, Column (3) conditions on $p_2 \geq 5$. Cluster-robust standard errors in parentheses. All specifications include order-effect dummies. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Replication of Table 3: Average θ for different ranges of module 2 price p_2

	(1)	(2)	(3)
	Standard	Triple	Pooled
High p_2 bin	0.145 (0.168)	0.191*** (0.063)	0.197*** (0.063)
Middle p_2 bin	-0.016 (0.132)	0.135*** (0.047)	0.142*** (0.046)
Std. tax cons.	0.131 (0.126)		0.044 (0.087)
Triple tax cons.		0.224*** (0.045)	0.217*** (0.045)
Observations	54522	54987	82028

Notes: This table replicates columns (1)-(3) of Table 3. Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm, and Column (3) provides estimates for the pooled data. For column (3) we include two-sets of moment equations for each arm, and we use the two-step GMM estimator to construct the weighting matrix since the model is overidentified in this case. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Replication of table 4: Average θ by ability to correctly compute tax, financial sophistication, and household income quartiles

	(1)	(2)	(3)	(4)
Compute tax correctly	0.218** (0.087)			0.168* (0.088)
Financially sophisticated		0.243*** (0.078)		0.160** (0.081)
Inc. quartile 2			0.038 (0.109)	0.013 (0.110)
Inc. quartile 3			0.234** (0.113)	0.180 (0.114)
Inc. quartile 4			0.338*** (0.111)	0.272** (0.115)
Std. tax cons.	0.105 (0.115)	0.124 (0.102)	0.091 (0.115)	-0.068 (0.133)
Triple tax cons.	0.318*** (0.074)	0.358*** (0.052)	0.320*** (0.080)	0.161 (0.103)
Observations	58005	58517	58517	58005

Notes: This table shows how average θ varies by ability to compute tax, financial literacy, and income quartiles. Data is pooled from both the standard and triple tax conditions.

Column (1) estimates a model with only computational ability as a covariate; column (2) estimates a model with only financial literacy as a covariate; column (3) estimates a model with only income quartiles as covariates; and column (4) estimates a model with all of three types of covariates.

Because we are using the pooled data, we include two-sets of moment equations for each arm, and we use the two-step GMM estimator to approximate the efficient weighting matrix for the over-identified model. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on module 2 price (p_2) being greater than 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Replication of Table 5: Predictiveness of self-classifying survey responses

	(1)	(2)
	C=0x,1x	C=0x,3x
“Yes” average θ	0.621*** (0.229)	0.627*** (0.081)
“A little” average θ	0.297*** (0.097)	0.412*** (0.040)
“No” average θ	-0.241* (0.126)	-0.025 (0.045)
Observations	54522	54987

Notes: Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.10 Replication of Main Results with OLS Regressions

Table A10: Replication of Table 2: Estimates of θ by experimental arm

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

	(1)	(2)	(3)
	All	$p_2 \geq 1$	$p_2 \geq 5$
Tax \times standard	0.258** (0.105)	0.254*** (0.088)	0.203** (0.082)
Tax \times triple	0.488*** (0.043)	0.479*** (0.038)	0.532*** (0.039)
Observations	60000	58517	32816

Notes: Dependent variable is $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$. Column (1) uses all data, Column (2) conditions on $p_2 \geq 1$, Column (3) conditions on $p_2 \geq 5$. Cluster-robust standard errors in parentheses. All specifications include order-effect dummies. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Replication of Table 3: Average θ for different ranges of module 2 price p_2

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

	(1)	(2)	(3)
	Standard	Triple	Pooled
Tax \times high bin	0.053 (0.164)	0.136* (0.070)	0.134* (0.071)
Tax \times middle bin	-0.128 (0.133)	0.103** (0.051)	0.110** (0.051)
Tax \times standard arm	0.288** (0.133)		0.179* (0.093)
Tax \times triple arm		0.414*** (0.052)	0.411*** (0.051)
Observations	40690	39397	58517

Notes: This table replicates columns (1)-(3) of Table 3. Dependent variable is $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$. Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm, and Column (3) provides estimates for the pooled data. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Replication of Table 4: Average θ by ability to correctly compute tax, financial sophistication, and household income quartiles

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

	(1)	(2)	(3)	(4)
Tax \times Quartile 2			0.039 (0.109)	0.029 (0.111)
Tax \times Quartile 3			0.272** (0.113)	0.242** (0.115)
Tax \times Quartile 4			0.442*** (0.110)	0.409*** (0.114)
Compute tax correctly	-0.009 (0.009)			-0.010 (0.009)
Financial literacy		0.004 (0.008)		0.008 (0.008)
Tax \times standard	0.117 (0.110)	0.162* (0.096)	0.061 (0.112)	-0.069 (0.132)
Tax \times triple	0.330*** (0.075)	0.393*** (0.053)	0.289*** (0.082)	0.153 (0.107)
Observations	58005	58517	58517	58005

Notes: Dependent variable is $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition module 2 price (p_2) being greater than or equal to 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Replication of Table 5: Predictiveness of self-classifying survey responses

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

	(1) Standard	(2) Triple
Tax $\times R = H$	1.068*** (0.281)	0.922*** (0.096)
Tax $\times R = M$	0.438*** (0.104)	0.627*** (0.047)
Tax $\times R = L$	-0.181 (0.123)	0.047 (0.052)
Observations	40690	39397

Notes: Dependent variable is $y_{ik} = \log(p_2^{ik}) - \log(p_1^{ik})$. Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.11 Replication of Main Results Excluding study participants not understanding the BDM Mechanism

Table A14: Replication of Table 2: method of moments estimates of $\bar{\theta}$ by condition

	(1)	(2)	(3)
	All	$p_2 \geq 1$	$p_2 \geq 5$
Std. tax avg. θ	0.283** (0.126)	0.276*** (0.103)	0.263*** (0.094)
Triple tax avg. θ	0.525*** (0.052)	0.512*** (0.044)	0.595*** (0.046)
Observations	46580	45411	25664

Notes: Column (1) uses all data, Column (2) conditions on $p_2 \geq 1$, Column (3) conditions on $p_2 \geq 5$. Cluster-robust standard errors in parentheses. All specifications include order-effect dummies. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Replication of Table 3: Average θ for different ranges of module 2 price p_2

Model: $\bar{\theta} = \alpha_0 + \alpha_{p_2 \in [5,10]} \mathbf{1}_{p_2 \in [5,10]} + \alpha_{p_2 \geq 10} \mathbf{1}_{p_2 \geq 10}$

	(1)	(2)	(3)
	Standard	Triple	Pooled
High p_2 bin	0.251 (0.206)	0.230*** (0.083)	0.257*** (0.082)
Middle p_2 bin	-0.154 (0.155)	0.157** (0.061)	0.173*** (0.061)
Std. tax cons.	0.299* (0.158)		0.104 (0.105)
Triple tax cons.		0.409*** (0.061)	0.392*** (0.060)
Observations	31358	30382	45411

Notes: This table replicates columns (1)-(3) of Table 3. Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm, and Column (3) provides estimates for the pooled data. For column (3) we include two-sets of moment equations for each arm, and we use the two-step GMM estimator to construct the weighting matrix since the model is overidentified in this case. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Replication of Table 4: Average θ by ability to correctly compute tax, financial sophistication, and household income quartiles

	(1)	(2)	(3)	(4)
Compute tax correctly	0.222** (0.098)			0.172* (0.098)
Financially sophisticated		0.243*** (0.088)		0.165* (0.091)
Inc. quartile 2			0.080 (0.124)	0.056 (0.126)
Inc. quartile 3			0.292** (0.130)	0.231* (0.132)
Inc. quartile 4			0.394*** (0.126)	0.328** (0.130)
Std. tax cons.	0.120 (0.127)	0.145 (0.113)	0.074 (0.129)	-0.096 (0.149)
Triple tax cons.	0.347*** (0.084)	0.390*** (0.060)	0.313*** (0.093)	0.141 (0.118)
Observations	45074	45411	45411	45074

Notes: This table shows how average θ varies by ability to compute tax, financial literacy, and income quartiles. Data is pooled from both the standard and triple tax conditions.

Column (1) estimates a model with only computational ability as a covariate; column (2) estimates a model with only financial literacy as a covariate; column (3) estimates a model with only income quartiles as covariates; and column (4) estimates a model with all of three types of covariates.

Because we are using the pooled data, we include two-sets of moment equations for each arm, and we use the two-step GMM estimator to approximate the efficient weighting matrix for the over-identified model. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on module 2 price (p_2) being greater than 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Replication of Table 5: Predictiveness of self-classifying survey responses

	(1)	(2)
	C=0x,1x	C=0x,3x
“Yes” avg. θ	0.843*** (0.298)	0.984*** (0.113)
“A little” avg. θ	0.575*** (0.127)	0.660*** (0.054)
“No” avg. θ	-0.219* (0.131)	0.066 (0.064)
Observations	31358	30382

Notes: Column (1) provides estimates for the standard tax arm, Column (2) provides estimates for the triple tax arm. Cluster-robust standard errors (at the subject level) in parentheses. All specifications include order-effect dummies and condition on $p_2 \geq 1$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F Financial Literacy Questions

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

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

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

G 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

<p>Envision Home Microfiber Bath Mat with Memory Foam, 16 by 24-Inch, Espresso</p>	<p>\$10.82</p>	<p>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.</p>
<p>Carnation Home Fashions Hotel Collection 8-Gauge Vinyl Shower Curtain Liner with Metal Grommets, Monaco Blue</p>	<p>\$8.99</p>	<p>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 repellent 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</p>

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.